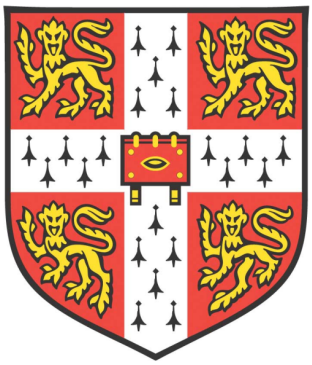
ROLE OF DDX RNA HELICASES IN CANCER CELLS



Ester Cannizzaro

St John's College

Department of Pathology

The Gurdon Institute

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

March 2018

Ester Cannizzaro

ROLE OF DDX RNA HELICASES IN CANCER CELLS

Abstract

DEAD-BOX (DDX) RNA helicases are a large family of proteins characterized by the presence of a DEAD/H (Asp-Glu-Ala-Asp/His) motif. Their main function is to unwind double stranded RNA to promote downstream molecular events. They are involved in virtually all steps of RNA metabolism such as transcription, translation, RNA export and degradation, ribosome biogenesis and pre-mRNA splicing. The aim of my work was to investigate the functions of human DDX3X and DDX54 RNA helicases, particularly in *in vitro* cancer models.

To provide insight into their molecular functions, I identified RNAs bound by DDX3X and DDX54 in breast cancer (MCF7) cells by performing iCLIP experiments. This generated two very distinct RNA binding profiles: DDX3X preferentially bound exonic regions of mRNAs encoding translational factors, whilst DDX54 preferentially bound non-coding RNAs and intronic regions of mRNAs encoding nuclear proteins. Further bioinformatic analysis identified a few discrete binding motifs within DDX3X target RNAs. One of these, within the human *JUND* transcript, was validated as a DDX3X binding site using electrophoretic mobility shift assays.

Notably, the levels of proteins encoded by mRNAs bound by DDX3X were altered following knockdown of the helicase. These data highlight the importance of DDX3X in maintaining appropriate levels of certain

proteins, which in turn may explain at least some of the changes in phenotype observed upon DDX3X knockdown. In this regard, I showed that knocking down DDX3X or DDX54 in MCF7 cells slowed cell proliferation by inducing a G1/S phase arrest. Furthermore, a CRISPR/Cas9 dropout screen in a leukaemia cell line (MLL-AF9) showed that both helicases are essential for growth of these cells. However, loss of DDX3X or DDX54 had little effect on proliferation of immortalized NIH3T3 cells indicating that their loss is not generally lethal. In subsequent CRISPR/Cas9 dropout screens in various other cancer cell lines, including some derived from solid tumours, DDX54 was found to be essential for growth of all cell lines. In contrast, DDX3X was required for proliferation of only a subset of these. Focusing on DDX3X, I identified that the integrity of the helicase's RNA binding domains is essential for growth and cell cycle progression of an acute myeloid leukaemia cell line (OCI-AML3).

Overall, my findings shed mechanistic insight upon the role of DDX RNA helicases in cancer and identify DDX3X and DDX54 as potential targets for therapeutic intervention in certain cancers.

ACKNOWLEDGEMENTS

They say the main point of doing a PhD is to learn. I'm not sure I can summarize the amount and quality of things I have learned during the past years.
So first of all, I want to be grateful to my PhD.
For giving me the chance to understand undisclosed things and to face the challenge of choosing the paths to follow, for letting me down, for making me happy, for teaching me so much about science, life, people and myself.

Thanks to Tony, for giving me the chance to work in a stimulating and warm environment such as the TK lab.
Thanks to Andy, for helping me until the very last day, and to Helena, who supported me in the hardest times. Thanks to Mark, for being a friend and a mentor, even from Down Under!
Thank you to all my brilliant lab colleagues and friends, for their invaluable scientific help and great time, great dinners, great laughs we shared.

Thanks to England! Gloomy, special and welcoming.

Thank you so much to my many friends in Cambridge and to the ones far away.
During these years, I have met so many people from many different parts of the world and this made me so much richer.
A little mention goes to my Italian theatre crew. The experience of acting has taught me the greatest thing of all: enthusiasm is the basic requirement for anything you will ever succeed in.
The little bright side of missing Italy is you can share the feeling with so many people. Sometimes I barely felt I was abroad, so thank you to all the so many Italian mates I met on the way, for their friendship and that innate bond we just can't deny.

Nevertheless, I was so lucky to have my old friends around me and share with them the good and bad of these years. Thanks to Elena, Silvia and Edo for keeping me alive in so many ways (particularly with good meals).

Thanks to Mario for being on my side. Your love and persistent help and support got me to this day. Thank you for being there holding me. Thank you for being the balance when I lose it.

And finally, thanks to my Mum and Dad and to Alfredo. As all pages of my life, these as well belong to you. None of this would have been possible without your presence in all of my days. Although nothing will ever pay back the sacrifice we made living apart, I hope my small achievement can make you proud and happy.

Dedico questa tesi ai miei genitori e a mio fratello.
Come tutte le pagine della mia vita, anche queste appartengono a voi. Niente di tutto ciò sarebbe stato possibile senza la vostra presenza in tutti i miei giorni. Anche se nulla potrà mai ripagare il sacrificio che abbiamo fatto vivendo lontani, spero che il mio piccolo traguardo possa rendervi orgogliosi e felici.

DECLARATION

The work described in this thesis was carried out by the author at the Gurdon Institute, under the supervision of Professor Tony Kouzarides.

The work described here is original except where indicated by reference. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the word limit of 60,000 words as prescribed for Degree Committee for the Faculty of Biology.

Ester Cannizzaro

Signed: Ester Cannizzaro

Date: 16th MARCH 2018

TABLE OF ACKNOWLEDGEMENT OF ASSISTANCE

Data obtained from a technical service provider: Gurdon Institute sequencing facility CRUK Cambridge Institute Genomics Core
Bioinformatics data analysis: Dr Namshik Han Dr Sam Robson Dr Tommaso Leonardi

TABLE OF CONTENTS

TABLE OF FIGURES

Chapter 1 - Introduction

By integrating intrinsic and external information, cells coordinate multiple regulatory mechanisms to modulate gene expression. Control of gene expression occurs at two distinct levels: transcriptional and post-transcriptional. Significant progress has been made in understanding how gene transcription is regulated, however much less is known concerning post-transcriptional control of RNA molecules. In eukaryotes, transcription and translation are two uncoupled processes which take place in different subcellular compartments: the nucleus and the cytoplasm respectively. This determines a complex network of newly transcribed pre-mRNA processing events, all of which constitute different steps of gene expression regulation. Post-transcriptional regulation includes events such as maturation, export, localization, turnover and translation of mRNAs. The regulation of each of these events determines the appropriate time and place at which a given protein is produced [1]. The overall aim of the work described in this thesis is to provide insight into the mechanisms underlying post-transcriptional regulation of gene expression.

As soon as pre-mRNAs emerge from their sites of transcription they are targeted by a multiplicity of RNA binding proteins (RBPs) which mediate pre-mRNA processing reactions. The central role of RBPs in controlling gene expression at the post-transcriptional level, highlights their importance to the cell. By regulating post-transcriptional events, RBPs are instrumental in determining biological processes ranging from embryo development to differentiation [2-5].

RBPs bind primary transcripts and mediate processing of their 5' and 3' ends through addition of a cap structure and a poly(A) tail, respectively. They are also essential for assembly of the spliceosome, for intron selection and removal, and the joining of appropriate exons. Different

RBPs bind RNA within the spliceosome to determine the sequence identity of the mature RNA. Transport of mature mRNAs to the cytoplasm for translation is also mediated by RBPs containing nuclear export signals (NES). In some cases, RBPs contain nuclear localization signals (NLS) as well and function as shuttles between the nucleus and the cytoplasm. Once in the cytoplasm, an assembly of cytoplasmic mRNA and RBPs (messenger ribonucleoproteins - mRNPs) bind to motor proteins. This event localizes the transcript at the relevant subcellular region for translation. mRNAs are then assembled with translational initiation factors and ribosomes and translated or, alternatively, stored in granule structures. mRNAs can either exit the storage bodies due to a translational activation signal or they can be eventually degraded [6-8] (Figure1.1). Thus, RBPs are critically important regulators of every post-transcriptional step of gene expression.

In the next sections, I describe the main steps of RNA metabolism from the perspective of RNA binding proteins (RBPs) that mediate the processes involved. I will discuss the consequences of dysregulation of these processes, many of which have been shown to lead to cancer. Focused sections are dedicated to the description of RNA helicases and, in particular, two members of the DEAD-box RNA helicase subfamily, DDX54 and DDX3X.

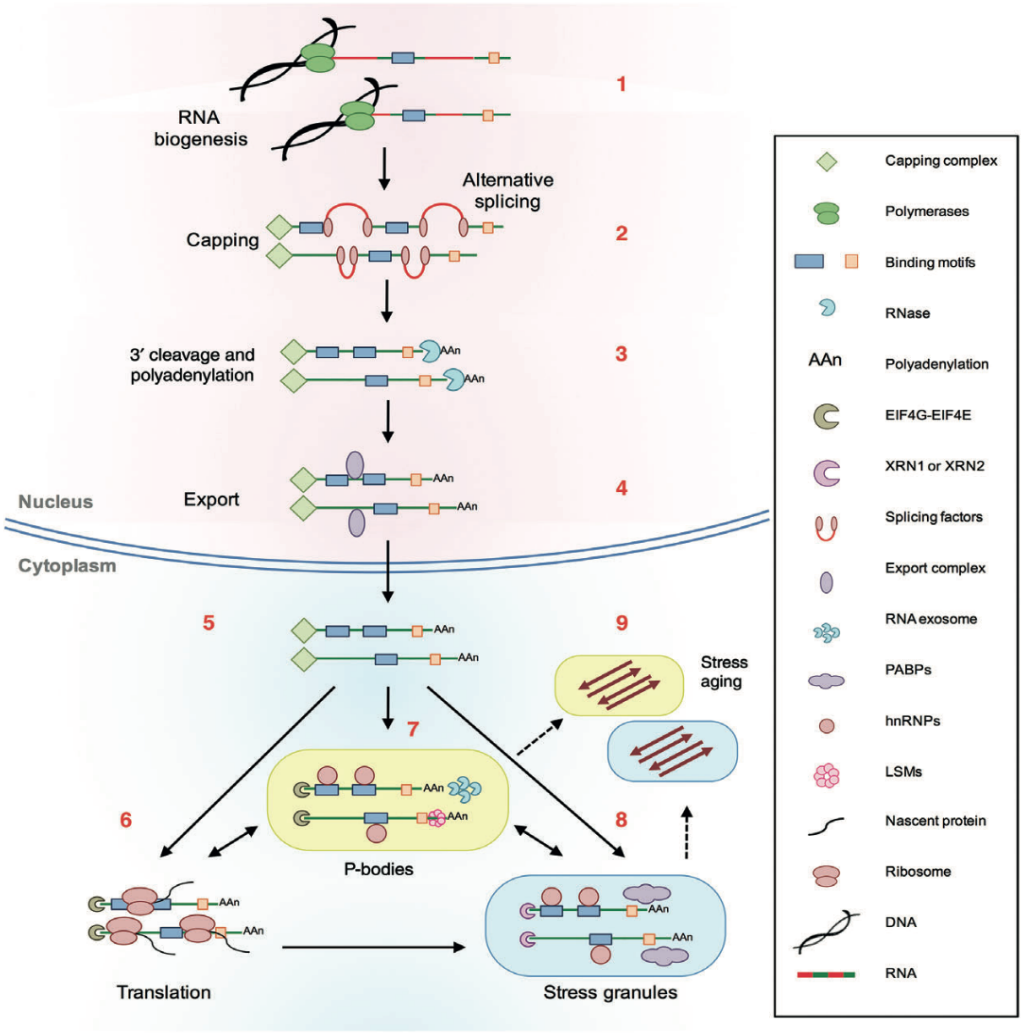


Figure 1.1. RNA-binding proteins (RBPs) and RNA metabolism.
RBPs are involved in every step of RNA metabolism from transcription to degradation. As soon as RNA is transcribed from DNA (pre-mRNA) (1) it becomes the target of maturation protein complexes that recognize sequence motifs within the transcript and mediate capping, splicing and polyadenylation events (2-3). The resulting mature transcript (mRNA) is then ready to be exported by means of other RBPs through the nuclear pores (4). Once in the cytoplasm (5), mRNAs are translated into proteins (6), stored (7-8) or degraded (9). Different RBP complexes mediate these events. (LSMs proteins are RNA chaperone ring-shaped proteins)[9].

1.1 mRNA metabolism: the long path of coding RNAs together with their protein partners

1.1.1 pre-mRNA nuclear processing

As transcription proceeds, nascent transcripts become available to be bound by proteins for pre-mRNA maturation. Newly transcribed RNAs undergo a series of modifications within the nucleus to produce a mature transcript that is ready for export to the cytoplasm and translation. The main and immediate pre-mRNA processing events are acquisition of a cap structure at the 5′-end, splicing out of introns and generation of poly(A) tail at the 3′-end. These reactions happen simultaneously with transcription [10] as many processing RBPs (such as capping factors (CFs), and 3′-end processing/termination factors) are recruited to the carboxy-terminal domain (CTD) of transcribing RNA polymerase II complexes. As soon as the first 20–30 nucleotides of the pre-mRNA have been synthesized, a cap structure is added at its 5′-end. This process requires the action of an RNA 5-triphosphatase (RTP) that hydrolyses the triphosphate of the first nucleotide to a diphosphate allowing a guanylyltransferase (GT) to catalyze the fusion of a GMP molecule. A methyltransferase ultimately methylates the N7 position of the transferred GMP. This cap structure is recognized by the cap binding complex (CBC) which plays a major role in the stabilization of the mRNA, avoiding 5′-3′ exonucleases activity [11, 12].

Nascent transcripts are susceptible to additional modifications, such as adenosines to inosines (A-to-I) RNA editing. In this process RNA deaminases of the ADAR family (ADAR1, ADAR2 and ADAR3) convert specific adenosines to inosines (A-to-I) through hydrolytic deamination. The pre-mRNA substrate of editing is a double-stranded RNA (dsRNA) usually formed by imperfect base-pairing between an exon that contains the adenosine to be edited and an intronic non-coding element [13] [14]. The resulting inosine bases are read as guanosines by the translation

machinery, which can result in incorporation of an amino acid different to the one encoded by the genome. Aberrations of ADAR-dependent RNA editing are linked to cancer progression and growth: hyperactivation of ADAR1 has been linked to progression of breast and hepatocellular cancers, as well as to haematopoietic disorders such as myeloproliferative neoplasms (MPN) and chronic myeloid leukaemia (CML) [15-17].

The addition of a poly(A) tail to the 3'-end of an RNA requires assembly of a multimeric and multifunctional RBP complex onto the pre-mRNA. Endonucleolytic cleavage occurs 10 – 30 nucleotides downstream of a highly conserved signal sequence (the AAUAAA hexamer) that is recognized by the cleavage and polyadenylation specificity factor (CPSF) [18, 19]. The addition of a poly(A) tract at the 3'-end of the RNA is then performed by Poly(A) polymerase (PAP) which is loaded onto the pre-mRNA together with the CPSF [20]. Poly(A) binding protein (PABP) binds the emerging poly(A) tail and in turn enhances the processivity of the poly(A) polymerase [21] [12]. Similarly to the 5′-cap structure, polyadenylation of an RNA affects its nuclear transport, translation efficiency and stability [12, 22]. Studies have highlighted the importance of the poly(A) tail length in gene expression control. A large proportion of genes contain more than one polyadenylation site created as part of alternative polyadenylation processes. This phenomenon, generates mRNA isoforms that differ either in their coding sequence or in the length of their 3′ untranslated regions (UTRs). Interestingly, cancer cells express mRNA isoforms with shorter 3′-UTRs [23, 24]. Since the 3′-UTR is a target of predominantly negative regulatory factors, this generally results in production of excess of proteins. In gliomas, this phenomenon promotes cancer phenotype by causing an accumulation of transcripts encoding proteins involved in cellular proliferation [25].

The next step in pre-mRNA maturation is RNA splicing. More than 90% of human coding genes contain multiple exons. The pre-mRNAs transcribed from those genes undergo splicing to generate mature mRNAs ready to be translated [26]. Splicing is carried out by a large macromolecular

machinery known as the spliceosome which consists of five small nuclear ribonucleoprotein complexes (snRNPs - U1, U2, U4, U5 and U6) responsible for the excision of non-coding introns from the pre-mRNA and joining of exons. The premature transcripts contain splicing recognition sequences at exon - intron boundaries (the 5′ splice site) and at intron – exon boundaries (the 3′ splice site). These are essential for targeting of the RNA by proteins within the spliceosome. Assembly of the splicing machinery initially requires interaction of U1 and U2 snRNPs with the 5' splice and 3' splice sites, which then recruits the U4/U6/U5 tri-snRNP to start the process [27] (Figure 1.2 A). Splicing can be constitutive or alternative depending on whether the splice sites are always or only sporadically recognized and spliced into mature mRNA. Although the same machinery is involved in both cases, different trans-acting RNA binding proteins govern recruitment of the spliceosome to the relevant alternative splice sites [28]. Among these RBPs include the serine/arginine-rich proteins (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNPs) [29]. In general, SR proteins promote splicing (exon inclusion) whilst hnRNPs repress it (promote exon skipping), although SR proteins can also act as repressors in a context-dependent manner [30] [31] [32](Figure 1.2 A). Notably, alternative splicing is observed on almost all human genes that contain multiple exons and this obviously expands the mature transcriptome as well as the subsequent proteome [33].

SR proteins are characterized by the presence of a C-terminal domain enriched in arginine (R) and serine (S) amino acids (RS domain) that mediate protein-protein interactions determining the recruitment of the spliceosome. SR proteins all possess at least one N-terminal RNA recognition domain (RRM domain), responsible for protein-RNA interaction[30]. RNA binding domains are discussed in the next section (Section 1.2) (Figure 1.2 B).

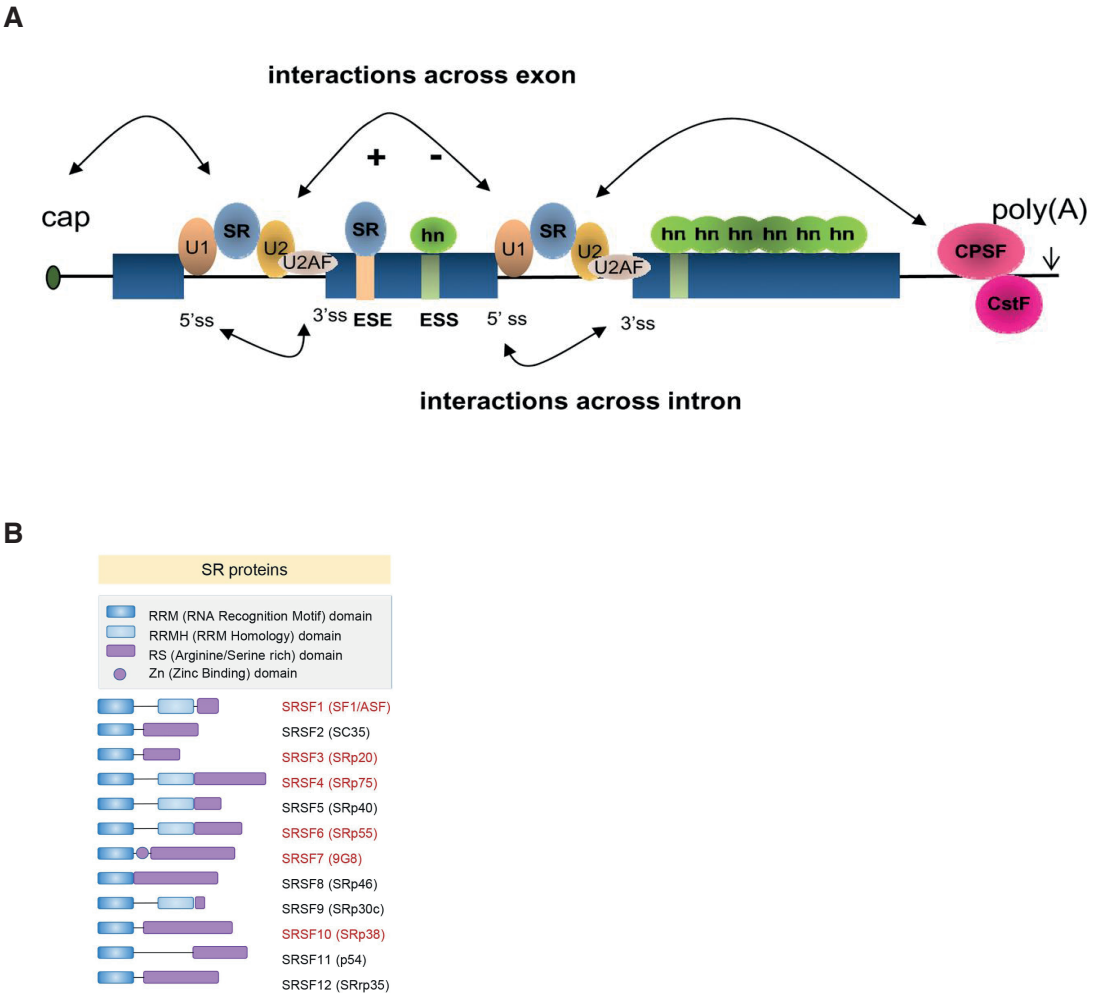


Figure 1.2. Constitutive and alternative splicing are performed by different RBPs.

(A) Arrows indicate the protein-RNA interactions across exons, between the cap and poly A synthesis factors (CPSF/CstF) (upstream) and the introns (below). U1, U2 and U2AF snRNPs bind constitutive splicing recognition sequences at the exon/intron boundaries. SR proteins bind ESE positive (+) sites and hnRNP proteins bind ESS (-) negative sites of alternative splicing. (hnRNPs; green shaded ovals; SR: blue ovals; U1: pink shaded oval; U2 snRNP :yellow shaded oval; U2AF :beige shaded oval; ESE :light orange box; ESS : green box; 5′cap: dark green small oval; 3′-polyA machinery CPSF/CstF: dark pink ovals; Exons are shown as blue shaded boxes, and introns and 5′- and 3′-untranslated regions are shown as black lines) (exonic splicing enhancers-ESE; exonic splicing suppressors-ESS) [34] (B). SR proteins are responsible for alternative splicing. The panel shows the main known SR proteins and their structural domains (RRM, RRMH, RS, and Zn) [30].

Splicing regulation is essential for cell growth, differentiation and development. As a consequence, alterations of the splicing process play a role in cancer development. Spliceosomal mutations were initially discovered in haematological malignancies, including myelodysplastic syndromes (MDS) and chronic lymphocytic leukaemia (CLL) [35] [36, 37]. More recently they have also been identified in several solid tumour types [38-40]. Furthermore, both hnRNPs and SRs are mutated or deregulated in different cancers, including lung, colon and breast cancer and haematological malignancies, where they act as either onco-proteins or tumour suppressors [41] [35, 37, 40, 42, 43] .

Upon excisions of introns and conjunction of exons, mature spliced mRNAs are bound by the exon junction complex (EJC). The EJC is a multiprotein complex organized around a central core of four proteins: eIF4A3 (eukaryotic initiation factor 4A3), MAGOH, Y14 (also known as RNA-binding motif 8A) and MLN51 (metastatic lymph node 51) [44]. The EJC is loaded onto mRNAs without sequence specificity. However, it does bind to a conserved position 24 nucleotides upstream of the spliced junctions [45]. The EJCs remain bound to the mRNA during its transport to the cytoplasm and until the translational initiation complex assembly causes it to dissociate from the mRNA [46]. During this time, the protein composition of the the EJC changes with different proteins associating or dissociating from it. These proteins are involved in different processes including splicing, transport, translation and non-sense mediated decay (NMD)[47-49]. This classifies the EJC as a bridging complex containing RNA binding proteins determining continuity between nuclear splicing and cytoplasmic processes. SR splicing factors 1, 3 and 7 (SRSF1, SRSF3, SRSF7) are associated with the EJC in both the nucleus and the cytoplasm [50]. The EJC also associates with the mRNP of the mRNA nuclear export factor TAP-p15, promoting passage of the newly matured mRNA through the nuclear pore towards the cytoplasm [47, 51]. Association of EJC with mature mRNA ready for nuclear export is

considered a mechanism of mRNA quality control that ensures only completely processed transcript can exit the nucleus [52, 53].

1.1.2 mRNA export

mRNA export is a complex and dynamic process involving many different RBPs. The complexes composed by the target mRNA and translocating proteins are formed in the nucleus, transported through the nuclear pore and then released into the cytoplasm where the mRNA contacts more RBPs in order to be translated, degraded or temporarily stored [54].

The two main pathways that mediate mRNA export across the nuclear pore complex (NPC) involve the NXF1-NXT1 (or TAP-p15) heterodimer and the exportin CRM1 (or XPO1). NXF1 is an RNA binding protein which assembles with the TRanscription-EXport complex 1 (TREX-1) to efficiently export mRNAs [55]. CRM1, is an export factor which does not possess intrinsic RNA binding domains and therefore needs an adaptor RBP in order to export mRNAs. NXF1-TREX1 and CRM1 are export receptors and they interact with the nucleoporins within the nuclear pore complex. This allows them to mediate translocation of mRNPs [56]. It is intuitive to predict that mRNA export is a vital step in regulating gene expression and therefore its mis-regulation may contribute to the development of cancer [57]. Indeed, deregulation of many components of the mRNP export machinery is correlated with different cancers. For example, CRM1 expression is elevated in gliomas, cervical and pancreatic cancers [58-60]. In addition, expression of components of the export complex TREX was found altered in oral, liver, pancreatic, lung, ovarian and colon cancers [61-64].

1.1.3 mRNA cytoplasmic fates

Once in the cytoplasm, the EJC bound to mRNPs is displaced by assembly of the translational initiation complex [46]. Persistence of the EJC on an mRNA indicates the presence of premature stop codons located up-stream of the EJC position [65]. Indeed, the best characterised function of EJC is its quality control of translation involving the nonsense mediated decay (NMD) pathway [66]. Normally, all EJCs are removed from transcripts after the first round of translation. When a transcript harbours premature translation termination codons, the presence of at least one EJC downstream of the stalled ribosome triggers NMD of the mRNA. Interestingly, since the EJC is loaded onto mRNAs within the nucleus during splicing, this suggests that mRNAs retaining premature termination codons could be marked for degradation even before getting to the cytoplasm [66].

The NMD pathway quality controls transcripts to ensure that faulty transcripts are destroyed to prevent accumulation of truncated proteins [67]. Assembly of an NMD-activating complex on a premature stop codon-containing mRNA requires the progressive recruitment of other RBPs along with the UPFs proteins, such as exonucleases and de-adenylases at the 3' end, and de-capping factors at the 5' end (de-capping enhancer Pro-rich nuclear receptor co-activator 2 -PNRC2). It is unknown whether all of these events occur simultaneously or sequentially [68] [69] [70] [71]. Cancer can be triggered by non-sense mutations in tumour suppressor genes. The NMD pathway has shown its relevance in favouring these tumours by mediating degradation of premature termination codon-containing mRNAs encoding tumour suppressors, such as TP53. [72].

By engaging protein-protein interaction with the cytoskeletal molecular motor factors and binding mRNAs' 3′-UTRs at the same time, some RBPs can target transcripts to particular subcellular localizations before translation [73].

Translation of mRNAs is probably the post-transcriptional process that involves the highest number of RBPs in a dynamic exchange of

interactions. Translation can be divided into three phases: initiation, elongation and termination. Of all steps, translational initiation is so far, the most studied and best understood. Translational initiation is a complex multistep process and is the main target of translational control mechanisms (Figure 1.3). The first step in the initiation pathway is assembly the 43S pre-initiation complex (43S PIC). 43S PIC is a large multifactorial complex comprising the 40S ribosomal subunit, eukaryotic translation initiation factors (eIFs) eIF1, eIF1A, eIF3, eIF5 and the ternary complex (TC). The TC is formed by initiation factor 2, eIF2, the initiator methionyl-tRNA (tRNA ^{Met}) and GTP. The eIF4F complex, consisting of the mRNA 5′-cap-binding subunit (eIF4E), the large scaffolding protein (eIF4G) and the DEAD box RNA helicase (eIF4A), is responsible for recruiting 43S PIC onto the mRNA via interactions between eIF4E–mRNA-cap and eIF4G–eIF3. Through the ATP-dependent action of eIF4A, the eIF4F complex assembles on the 5'-cap of the mRNA and unwinds secondary structures found in the 5'-UTR. Other factors (eIF4G) bridge the complex at the 5' UTR to the poly-A tail at the 3' end by interacting with the poly(A) binding protein (PABP). This causes mRNA circularization which stabilizes the transcript and stimulates translation.

The 43S PIC complex scans the transcript in a 5′ to 3′ direction in search of the initiation codon. When the 43S PIC complex encounters an AUG codon embedded within a favourable sequence context (the Kozak sequence), base pairing between the initiation codon and the initiator tRNA in the ternary complex anticodon takes place. This event then triggers GTP hydrolysis by eIF2, which causes release of eIFs and joining of the large 60S ribosomal subunit to the 40S Met-tRNA-mRNA complex. Formation of a translation-competent 80S ribosome ends the translational initiation process and triggers subsequent elongation (Figure 1.3).

During translational elongation, eEF1-aminoacyl-tRNA-GTP complexes bind to the ribosomal A site. Several steps involving codon-anticodon base pairing between the mRNA and the tRNA ensure that only the correct cognate tRNA is selected. Codon-anticodon base pairing induces eEF1-

GTPase activity which determines release of the aminoacyl tRNA into the A site, ready to form a peptide bond. A peptide bond is formed between the incoming amino acid and the peptidyl tRNA in the P ribosome site. The translocation of the ribosome is catalysed by eIF2 GTP hydrolysis, which determines the exit of the deacylated tRNA from the E site, the repositioning of the peptidyl tRNA in the P site, and the mRNA movement by three nucleotides to place the next codon into the A site. This cycle is repeated until a stop codon enters the A ribosomal site. At this stage, the completed polypeptide is released following the activity of GTP-ase release factors that hydrolyse the ester bond that links the polypeptide chain to the P site tRNA [74].

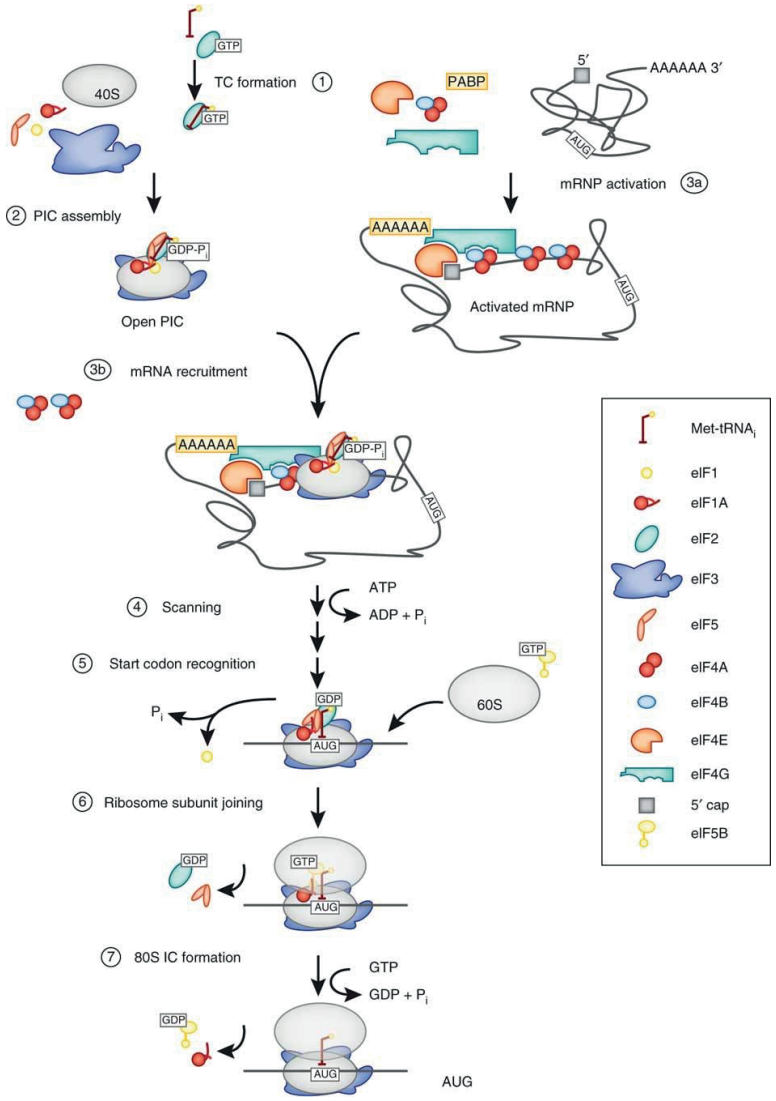


Figure 1.3. Translational initiation complex assembly.
The ternary complex (TC) assembles by the interaction of its components eIF2-GTP and the initiator tRNA (1). TC is then recruited to the 40S ribosome subunit through cooperative binding of eIFs 1, 1A, 3 and 5. The resulting assembly is the pre-initiation complex (PIC) (2). The mRNA is bound at the 5′end cap by the eIF4 factors and at the 3′ end polyA tail by PABP (3a), and recruited to the PIC (3b). The PIC complex scans the mRNA in a 5′ to 3′ direction up to the location of a start codon (AUG) (4). Upon start codon recognition, GTP hydrolysis releases eIF1 and arrest the scanning process (5) determining further dissociation of eIF2•GDP and eIF5 and allowing association of eIF5B-mediated joining of the 60S subunit (6). eIF5B-GTP hydrolysis and subsequent release leave the 80S initiation complex (IC) ready to mediate attachment of the first methionine (7) [75].

The translational apparatus responds to environmental changes and this ensures a cell responds to particular conditions such as metabolic stress or nutrient uptake. Regulation of translation is therefore very important in controlling cell growth and division. This tight control is accomplished by the induction of many pathways that ultimately lead to translational stimulation or repression [76]. In many cases signals can direct translation of selected proteins needed for the cell to respond to a specific condition. One pathway that exemplifies such a regulatory mechanism is the mTOR signalling pathway.

Mammalian target of rapamycin (mTOR) is a serine/ threonine kinase that exists in two functionally and structurally divergent complexes: mTOR complex 1 and 2 (mTORC1- mTORC2). mTORC1 affects cellular proliferation, growth, protein synthesis, metabolic programmes and autophagy [77, 78], whereas mTORC2 controls glucose and lipid metabolism as well as cytoskeletal organization [79].

The mTOR pathway integrates extracellular and intracellular growth signals, such as nutrients, growth factors or stress conditions, and links them to the translation process by regulating it at several levels. In fact, mTOR can mediate transcriptional regulation of ribosomal RNA (rRNA) and genes encoding ribosomal proteins, as well as control ribosomal proteins mRNA stability and regulate their translation. This leads to a coordinated response of the translational apparatus to external stimuli [80, 81].

mTORC1 directly regulates translation mainly by modulating assembly of the eIF4F complex. Activation of the mTOR pathway leads to phosphorylation of different effector proteins such as 4E-Binding proteins (4E-BPs). These are usually associated with eI4FE and inhibit its function. Upon mTOR phosphorylation, 4E-BPs dissociate from eIF4E which in turn allows eIF4E–eIF4G interaction and eIF4F complex assembly [82, 83]. The other main effectors of mTOR are the S6 kinases 1 and 2 (S6K1, S6K2). Activated mTORC1 phosphorylates and activates S6 kinases 1 and 2 [84, 85]. S6Ks in turn phosphorylate various substrates

among which are the ribosomal protein S6 [86] and the translation initiation factor eIF2 [87].

The mTOR pathway represents an example of transcript-specific regulation of translation. In particular, mTOR stimulates translation of mRNAs with a specific 5′ sequence, the 5′ terminal oligopyrimidine tract (TOPs) [88]. The 5′ TOP sequence is a stretch of 4–14 pyrimidines preceded by a C nucleotide. Most of the RNAs containing this sequence encode components of the translation machinery, including ribosomal proteins, eIF2 and poly(A)-binding protein (PABP) [89]. Ribosome profiling studies have shown that mTOR stimulates translation of mRNAs with TOPs and TOP-like pyrimidine-rich translational element (PRTE) motifs through phosphorylation and inactivation of 4E-BPs [90, 91]. Interestingly, it was later shown that upon stress conditions, such as hypoxia and nutrient deprivation, which inhibit the mTOR signalling, translation of TOP mRNAs was suppressed independently of 4E-BPs [92]. Different studies, proposed that S6Ks promote translation of the 5′-TOP-mRNAs via stimulation of RPS6 phosphorylation [93, 94]. However, it was subsequently shown that neither loss of S6Ks nor the phosphorylation status of RPS6 affected translation of 5′-TOP mRNAs [95]. These controversial observations suggest that more unknown effectors mediate regulation of translation of TOP mRNAs through mTORC1-dependent signalling [96].

Growing evidence shows that translation is frequently deregulated in cancer. Indeed, tumour cells have a high demand on protein synthesis to cope with their prolific growth and this requires higher levels of ribosomes and other translational components [97, 98]. The pathways that are often found to drive oncogenesis in many cancers, such as PI3K/ AKT/mTOR, RAS/MAPK, and Wnt/β-catenin, are typically upstream translational regulation pathways. This suggest that mis-regulation of translation is a hallmark of cancer and it can be achieved in different ways [85, 98]. One recurrent feature is overexpression of eIF4E, which is found in several

tumour types including head and neck, bladder, colon, breast, prostate, lung and blood cancers [99, 100].

Inhibitors targeting translational components or members of signalling pathways that activate translation have been considered for cancer therapeutic intervention. In particular, inhibition of mTOR has gained considerable attention. Rapamycin is a naturally occurring macrolide produced by *Streptomyces hygroscopicus* [101] that inhibits TOR kinases [102]. Rapamycin induces conformational changes that weaken the interaction between regulatory-associated protein of mTOR (RAPTOR) and mTOR [103], within the mTORC1 complex [104, 105].

Several rapamycin analogues (rapalogues) were generated to improve its pharmacodynamic properties, and these exert anti-neoplastic activity in cancer cell lines and mouse models [106]. Many of these molecules have been used in the clinic, although they have showed lower efficacy than hoped [107].

Inhibition of translation and consequent disruption of the translational machinery can induce formation of RNA-protein cytoplasmic aggregates in which untranslated RNAs are stored. The main types of such subcellular structures are the processing bodies (PBs) and the stress granules (SGs) [108, 109].

While PBs contain components of the mRNA decay machinery, SGs contain components of the translation initiation machinery. PBs form in unstressed cells during translational inhibition. In general, events that reduce the efficiency of translational initiation include disruption of interactions between poly(A)-binding protein 1 (PABP1) and the initiation complex and dissociation of translating ribosomes. These events lead to recruitment of de-adenylation and de-capping factors to the mRNA thereby inducing aggregation of mRNPs and decay proteins within PBs [108].

During stress response, eukaryotic translation initiation is reduced and this is associated with formation of stress granules triggered by aggregation of several RBPs [110]. The translational silencers TIA1 and

TIAR are recruited to the mRNP in response to stress, to promote assembly of stalled translation initiation complexes and formation of SGs [111]. Stress granules typically contain translational factors such as poly(A), 40S ribosomal subunits, eIF4E, eIF4G, eIF4A, eIF4B, Poly(A) binding protein (PABP), eIF3, and eIF2 [112]. When cells recover from stress, SGs disassemble and mRNAs re-enter translation [113] [114].

SGs do not mediate an impairment of global translation but rather of selected mRNAs. Usually, the mRNAs in SGs are those encoding proteins involved in cell cycle and housekeeping functions, the silencing of which will slow down cell growth in order to cope with stress [115] [116]. In contrast, mRNAs encoding stress-inducible proteins (HSP70 and HSP90) are excluded from SGs [117, 118]. This selectivity shows that aggregation of SGs is another means by which RBPs post transcriptionally regulate gene expression.

RNA granules affect cancer development mainly by favouring the adaptation of cancer cells to stress conditions of the tumour microenvironment [119]. Solid tumours are characterized by outgrowth due to rapidly proliferating cancer cells. This is often accompanied with instauration of deprivation of oxygen and nutrients which exposes the cancer cells to hypoxia and nutrient starvation [119]. To cope with such as environment, tumours often trigger a stress response which helps cells to adapt and survive to the changing conditions [120]. Formation of SGs in cancer cells has been shown to be part of this process. SGs can sequester and inactivate pro-apoptotic factors (such as RACK1 and TRAF2) [121, 122] and mRNAs encoding pro-apoptotic proteins that are induced by oxidative stress [123].

As mentioned earlier, mTOR is a sensor of the metabolic state of the cell. In cancer cells under conditions of nutrient starvation and stress, mTORC1 stimulates TIA1/TIAR dependent recruitment of 5'-TOP-containing mRNAs to stress granules, hindering their translation [124, 125]. Since this subclass of mRNAs predominantly encodes protein biosynthesis factors such as translation factors and ribosomal proteins,

this activity of mTORC1 represents an integration of the response to the stressful tumour microenvironment and an example of transcript specific translational regulation essential to cancer cell metabolism.

1.2 RNA binding proteins features

RNA binding proteins mediate formation of ribonucleoprotein complexes (RNPs) that lie at the core of post-transcriptional gene regulation processes. In the formation of RNPs, RNA acts as a scaffold guiding the protein's synergistic recruitment through their recognition of RNA sequence or structural features. Within RNPs, RNA is remodelled, matured, transported and rendered accessible for other RBPs to bind and for enzymatic reactions to take place.

It has been estimated that RBPs comprise 3–11 % of the entire proteome of bacteria, archaea and eukaryotes and they are significantly conserved throughout evolution [126, 127]. Many RBPs have been identified by the existence within of well-characterised RNA-binding domains (RBDs). The best-known RBD is the RNA recognition motif (RRM), composed of 80–90 amino acids, that is found in most RNA binding proteins and that can recognize short stretches of four to eight nucleotides without sequence specificity. Other known domains include K-homology (KH domain), RGG (Arg-Gly-Gly) box, Sm domain, DEAD/DEAH box helicase domain, zinc finger (ZnF), double stranded RNA-binding domain (dsRBD); cold-shock domain; Pumilio/ FBF (PUF or Pum-HD) domain; and the Piwi/Argonaute/ Zwille (PAZ) domain (Figure 1.4 A) [128]. Computational studies have revealed that eukaryotic genomes encode a large number of RBPs containing one or more of these domains, although recent proteomic studies of the RNA interactome in human cells have also identified hundreds of proteins previously unrelated to RNA metabolism and lacking any known RBDs. This suggests that the number of RNA binding proteins and potential RNA binding domains might be substantially higher than currently thought (Figure 1.4 B)[129-131].

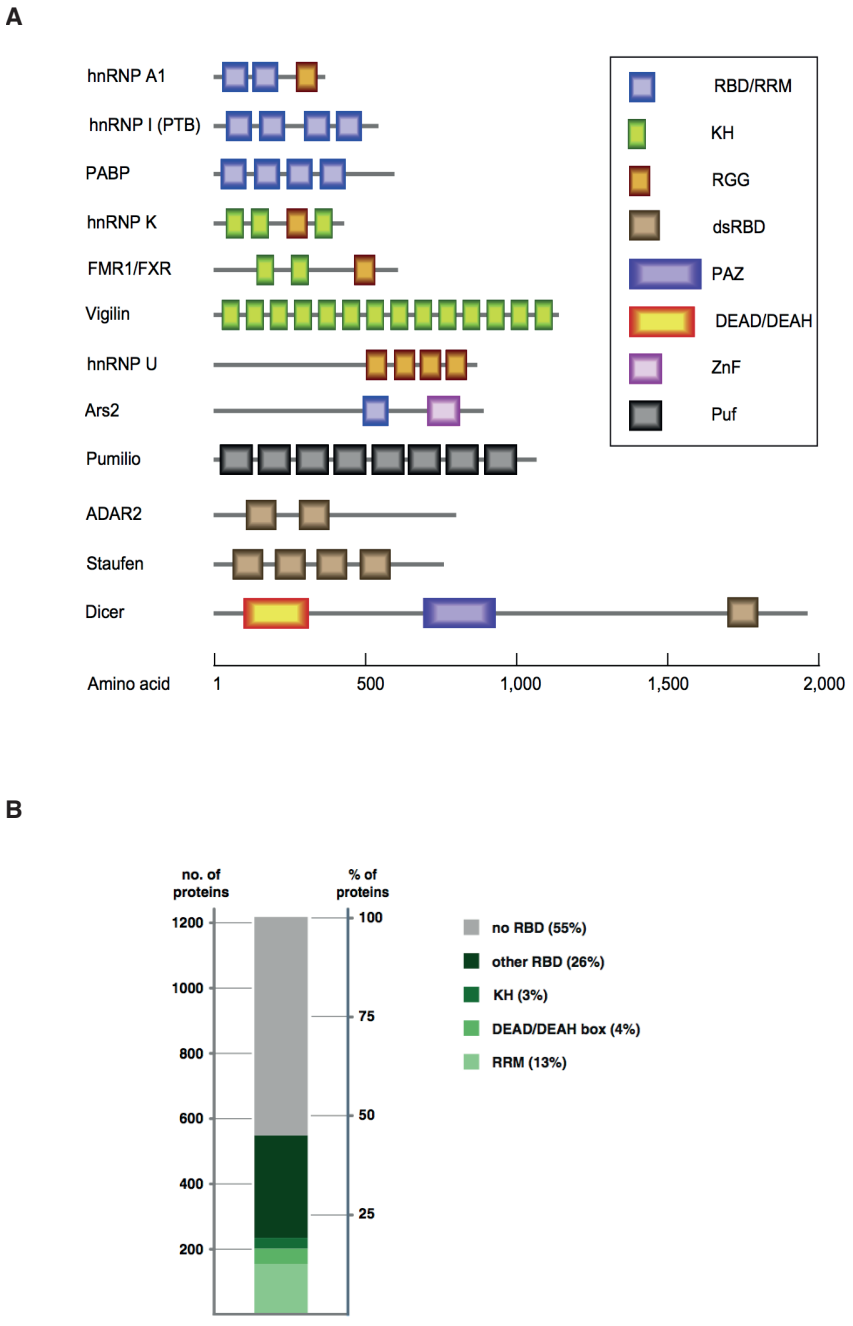


Figure 1.4. RNA binding domains in RBPs.
(A) shown are the most frequently found RNA-binding domains (RBD), including K-homology (KH) domain, RGG (Arg-Gly-Gly) box, double stranded RNA-binding domain (dsRBD), Piwi/ Argonaute/Zwille (PAZ) domain, RNA helicase DEAD/DEAH box, RNA-binding zinc finger (ZnF) and Puf RNA-binding repeats (Puf). All are presented as coloured boxes [1]. (B). RNA interactome capture in human cell lines. An RNA interactome experiment identified a total of 1218 proteins as RNA binders. Most proteins do not contain any known RBD (~55 %). The remaining proteins harbour the most common RBDs, the RNA recognition motif (RRM, ~13 %), DEAD/DEAD box helicase domain (~4 %) and the K homology domain (KH,~3 %) [130].

Most RBPs are composed of the combination of different RNA binding modules as well as auxiliary domains (Figure 1.4 A). RBPs are able to bind RNA with high affinity thanks to the synergistic binding of multiple modules, each of which exerts a weak interaction in isolation. Also, the spread distribution of RNA binding domains within the RBP structure, provides the protein with the ability to bind portions of RNA or regions distally rearranged by secondary RNA structures.
If the affinity of a RBP is mainly due to its modular composition and rearrangements, the specificity is also influenced by the ability of the proteins domain to also interact with other proteins. The protein-protein interactions can cause even small structural alterations that are transmitted to the RBD portions and favour the binding of specific RNAs. Therefore, the variation of interactions among RBPs determine their context-specific functional activity [132]. The alteration of their targets or the aberrant assortment of their protein interactions is therefore responsible for the role of RBPs in tissue-specific diseases, including cancer [133].

1.2.1 RNA helicases

RNA helicases are the largest class of RNA binding proteins. They are involved in all RNA-related processes and especially in translation initiation, ribosome biogenesis and pre-mRNA splicing. RNA helicase activity in these processes utilizes the energy from ATP hydrolysis to remodel RNA or RNA-protein complexes in order to separate RNA duplexes, displace proteins from RNA molecules, anneal RNA strands or act as RNA clamps to stabilize RNA folding intermediates [134, 135].

All RNA (and DNA) helicases can be identified by the presence of the typical Walker A and B sites for NTP binding and hydrolysis. A high number of proteins were structurally identified as helicases and classified into six super-families (SF1-6). Within each super-family, the helicases

share homology in terms of both sequence of motifs and structural arrangement. Proteins of SFs 1 and 2 are characterised by a helicase core formed by two structurally almost identical helicase domains, which contain all the conserved motifs. Helicases of SFs 3–6 form hexameric toroids. These proteins contain only one helicase domain that is distinct from helicase domains within enzymes of SFs 1 and 2.

Classification of a protein as a helicase has long being debated. This is because the presence of a helicase domain does not necessarily indicate that the protein possesses unwinding activity. In fact, while all of them are able to hydrolyse ATP in a nucleic-acid-dependent fashion, this activity is not necessarily coupled to RNA (or DNA) duplex unwinding. Therefore, the name helicase identifies the sequence features more than the enzymatic function. Moreover, the mechanism these enzymes use to unwind their double stranded nucleic acids is not necessarily common among super-families. Some helicases unwind by translocation, others translocate without unwinding or unwind without translocating [135-138]. In general, the only real common property of helicases, in terms of mechanistic and enzymatic activity, is their ability to bind nucleic acid in an ATP-dependent fashion.

SFs 1 and 2 contain the largest number of helicases. These proteins contain twelve conserved motifs that constitute the helicase core and confer similar activities to the different members. The conserved helicase core consists of two similar protein domains (domains 1 and 2) named recA-like domains after resemblance to the fold of the bacterial recA protein, which is a DNA recombination enzyme. Helicases within SFs 1 and 2 share almost identical folds and extensive structural similarity of the cores (Figure 1.5 A) [137].

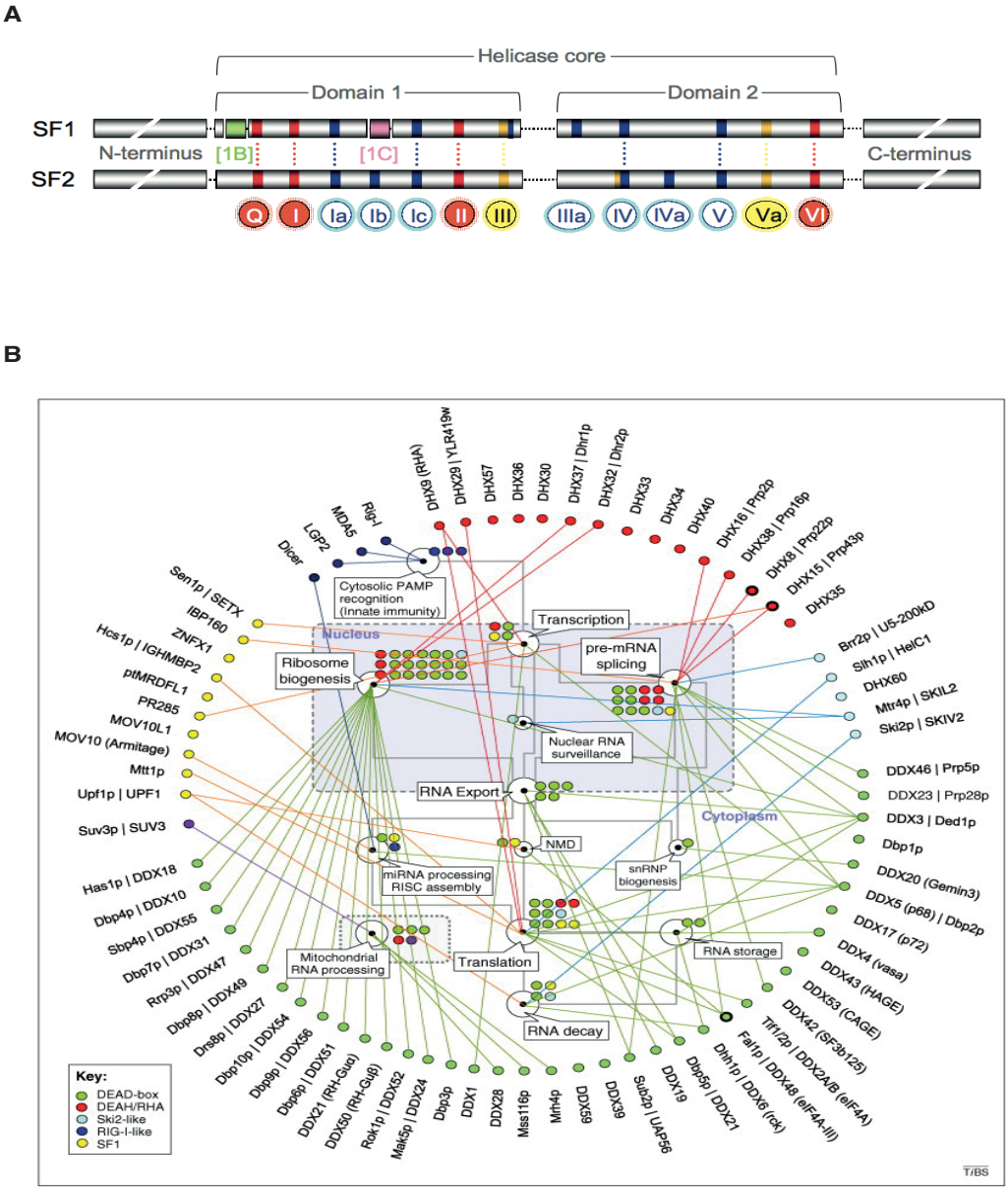


Figure 1.5. Structural features and functions of SF1 and SF2 RNA helicases.
(A) Comparison of the conserved sequence domains of SF1 and SF2 RNA helicases. 1B and 1C are SF1 specific inserts. (red: ATP binding and hydrolysis, blue: RNA binding, yellow: communication between ATP and RNA- binding sites). (B) Cellular roles of eukaryotic SF1 and SF2 RNA helicases. White circles and names indicate processes of eukaryotic RNA metabolism (NMD, nonsense mediated decay). Grey lines mark connections between processes. Coloured circles indicate the number and the subtype of individual RNA helicases involved in a given process. Yeast and human orthologues are grouped and color-coded according to their families (see legend at left lower corner). Coloured lined connect more processes for individual helicases and indicate involvement in one or more processes of RNA metabolism. Circles with bold lines emphasize the three RNA helicases (Prp22p, Prp43p, eIF4A-III) [139].

Within the helicase core, Motifs I (Walker A) and II (Walker B) are required for NTP/ATP binding and catalyze its hydrolysis, Motif III has been suggested to couple ATP binding and hydrolysis to helicase activity, Motif VI functions in ATP hydrolysis and Motifs Ia, Ib, IV, and V are responsible for binding of the substrate RNA. The helicase core is flanked by N- and C-terminal extensions. The composition of the extensions vary and often contains additional functional RNA-binding domains such as RRMs [137].

Helicases within SF1 are predominantly DNA helicases and are found in all kingdoms of life, including many viruses.

SF2 is the largest helicase superfamily, with members throughout the eukaryotic kingdom, most bacteria, archaea, and in several DNA and RNA viruses. SF2 is subdivided into at least 10 sub-families. Five of these (DEAD-box, DEAH/RHA, Ski2-like, RIG-I-like, and NS3/NPH-II family) contain mainly RNA helicases.

SFs 3–5 contain mainly RNA helicases that appear to have no eukaryotic counterparts [136].

1.3 DEAD-box RNA helicases

1.3.1 Structural features of DEAD-box RNA helicases

DEAD box proteins (DDX) belong to SF2. RNA helicases of the DEAD box family are present in all eukaryotic cells and in many bacteria and archaea (Figure 1.6 A) [140].

The DEAD-box is the largest family of RNA helicases with 37 members in humans. The motifs of their helicase core are conserved with similar positioning in all members of the protein family and they are flanked by variable auxiliary domains which are thought to be critical for the diverse individual functions of these enzymes [141]. Like the other SF2 members, the helicase core is made of two virtually identical recA–like domains

within which are distributed twelve characteristic sequence motifs. The DEAD (Asp-Glu-Ala-Asp) motif, which gives the name to the protein family, is located in Motif II. In the closed conformation of the protein monomer, the two helicase domains form a cleft which constitutes the ATP binding and hydrolysis site. Motifs I-II-VI and Q create the ATP contacts whereas, on the opposite side of the ATP binding pocket, motifs Ia-Ib-Ic-IV-IVa and V associate with RNA mainly through interactions with the sugar phosphate backbone. The Q-motif, whose name reflects a highly conserved glutamine residue, is located upstream of motif I [142] and was proposed to function as a sensor to determine the state of the bound ATP thereby regulating ATP binding and hydrolysis [143]. Motifs III and IVa are thought to mediate communication between the RNA binding site and the ATP binding and hydrolysis site, although this aspect is still not mechanistically clear [144].

The N- and C-terminal flanking domains of the helicase core vary in length and amino acid composition from one member of the family to another. Information about the structural features of these domains are limited, because most of the studies have focused on the isolated helicase core [145].

The conventional helicase mechanism is characterized by a directional movement of the protein along one nucleic acid strand of the duplex. This translocation needs the unwinding activity since progression of the protein along the nucleic acid strand displaces the complementary strand [146]. DEAD-box proteins are relatively poor "unwindases" and may be more appropriately considered as ATP-dependent RNA binding proteins. DEAD-box proteins use a non-conventional non-processive unwinding mechanism that is only effective on approximately two helical turns of duplex, which include fewer than 10 or 12 base pairs. This mechanism is known as local strand separation: DEAD box helicases load directly on a duplex RNA, assisted by a single strand extending either at the 3' or 5' end of the

double strand. These single strand regions have to be proximal but not necessarily covalently connected to the target double helix. Some DEAD-box helicases can perform unwinding as monomers but in many cases multiple protomers of a helicase are actually required and the single stranded extensions provide binding sites for additional monomers. Loading onto an RNA helix is not directional and can occur at either end of an RNA molecule, or internally, and on either strand. A DEAD-box helicase loaded on its target RNA duplex can then interact with other proteins for further molecular functions and begin to catalyse strand separation. This action requires binding of ATP but not its hydrolysis. As a consequence of conformational changes due to the RNA-binding, the helicase destabilises the helix locally over just a few base pairs, causing the immediate neighbour double strand to unwind, without any further enzymatic action. ATP and RNA are bound cooperatively by DEAD-box helicases and this determines the closure of the two recA-like domains and the unwinding of the two RNA strands. It is this closure that stimulates ATP hydrolysis. The ADP generated upon hydrolysis has reduced affinity for the DEAD box helicase cleft, causing release of both the tightly bound RNA strand and the protein. The enzyme is then recycled for unwinding of other RNA substrates [147, 148] (Figure 1.6 B).

In physiological mRNPs assemblies, RNAs rarely exceed one helical turn. Therefore, the mechanism of unwinding by local strand separation is well suited for the biological function DEAD box proteins exert in the cell [135], especially in order to promote rapid assembly/disassembly of different protein complexes on various transcripts [149].

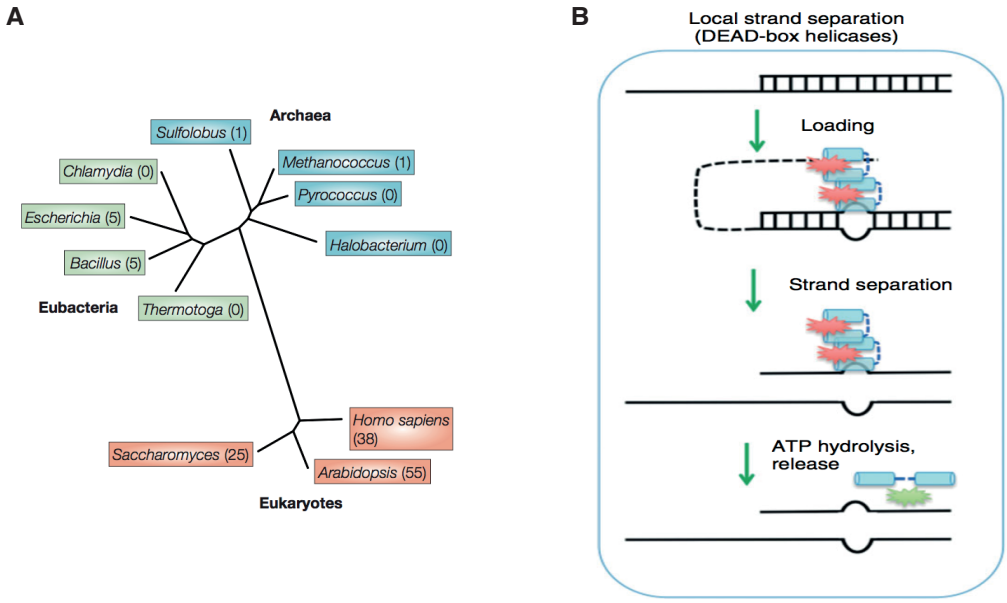


Figure 1.6. Conservation of DEAD box helicases across organisms and their mechanism of action.
(A) Representation of DEAD-box proteins from the three kingdoms of life. The phylogram was constructed by *Rocak et al* 2004 [140]. The numbers for *Homo sapiens*, *Arabidopsis thaliana* and *Saccharomyces cerevisiae* are from published reports [150-152] whereas the numbers for the remaining organisms were obtained by BLAST searches using the yeast eIF4A as reference sequence [140]. (B) Schematic representation of DEAD box RNA helicases unwinding mechanism by local strand separation. The RNA helicase in its ATP-bound (red star) closed conformation is loaded directly onto the RNA duplex without polarity or position preference. Loading of protomers is facilitated by single strand ends of the target RNA. Upon binding, the RNA duplex is locally destabilized, determining dissociation of the double strand. ATP hydrolysis determines release and recycling of the RNA helicase [146].

1.3.2 Biological functions of DEAD-box RNA helicases

DEAD-box RNA helicases are generally defined as RNA unwindases. However, their unwinding activity is only one of the several biochemical functions of these proteins. DDX helicases are able to unwind local secondary structures however, if the release of inorganic phosphate after ATP hydrolysis is inefficient the helicase can also function as an RNA clamp. As RNA clamps, DDXs can act as nucleation centres for the formation of RNA–protein complexes. Some can also mediate RNA-annealing activity alternatively to their RNA-unwinding activity, which gives DEAD-box proteins a wide ability to promote RNA rearrangements. DEAD-box RNA helicases can regulate RNA structures and RNA–RNA and protein–RNA interactions. They can also function as RNA chaperones by remodelling RNPs and modulating RNA–protein association/dissociation. DEAD-box proteins can remove proteins from RNA in an ATP-dependent reaction. Since DDXs do not translocate on their target RNAs, they can remodel RNA locally by displacing proteins from limited structured RNA [153].

As with all RNA binding proteins, DEAD box helicases are usually part of larger multi-component complexes, such as spliceosomes and ribosomes. It is intriguing to notice how, despite a very high structural conservation of their functional helicase core, these proteins perform very different functions in RNA metabolism, all of which require multi-step association/dissociation of large RNP complexes and modulation of RNA structures (Figure 1.7A).

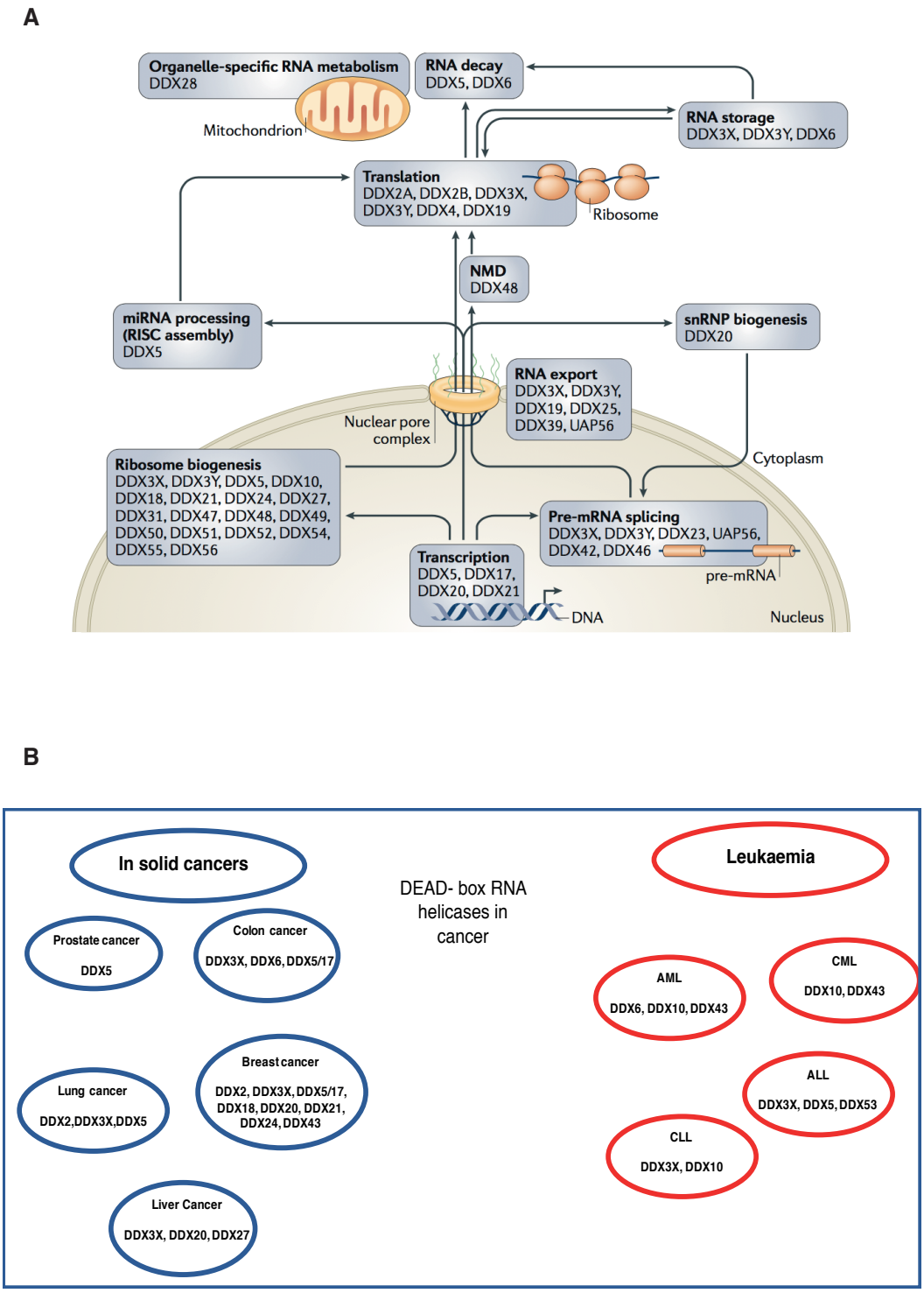


Figure 1.7. DEAD box RNA helicases perform many cellular functions and are associated with cancer.
(A) Human DDXs are involved in many steps of RNA metabolism as indicated. RNA helicases of the DEAD box family are widely involved in post-transcriptional regulation of gene expression (DDX, DEADbox; RISC, RNA induced silencing complex; snRNP, small nuclear ribonucleoprotein) [145]. (B) The figure groups DDXs reported to participate in the onset of various solid cancers or leukaemia subtypes. Adapted from [154].

DEAD-box proteins have been linked to transcription regulation, functioning as either co-activators or co-repressors (DDX1, DDX5, DDX17, DDX20, DDX2) through their interaction with key components of the transcriptional machinery [155]. Furthermore, many DEAD box helicases are involved in alternative splicing through interaction with either the exon junction complex (DDX48) [66] or the spliceosome (DDX41-DDX42) [156, 157].

Numerous DDX RNA helicases are also required for ribosome biogenesis (Figure 1.7A). This is a very dynamic process that requires processing of a ribosomal RNA (rRNA) precursor and further modification of the resulting 18S and 25S rRNAs that takes place in the nucleolus. The modifications (e.g. pseudouridylation and methylation) are performed by enzymes coupled to a class of small guide RNAs, called small nucleolar RNAs (snoRNA), that match the rRNA target sequence [140, 145]. It has been suggested that DEAD-box proteins are required for the unwinding of short duplexes of snoRNA–rRNA or rRNA–rRNA interactions to allow new and mutually exclusive RNA–RNA interactions to occur [158].

Different DEAD-box proteins mediate the export (DDX39A, DDX39B) [159] and translation (DDX48) [160] of selected mRNAs. DDXs play various roles at different steps of general translation: DDX2A and DDX2B (also known as eIFA1 and 2) constitute canonical translational initiation factors (Section 1.13), whereas other DDXs are required at later stages for recognition of stop codons and recruitment of translation termination complexes (e.g DDX19B) [161]. DDXs can also mediate mRNA degradation by stimulating de-capping enzymes (DDX6)[162].

The activities of these helicases on their target RNAs do not always strictly require their helicase function, which suggests that other functional domains are involved. Also, the outcome of DDX activity can either promote gene expression or repress it. By considering all of the

different functions these proteins can exert, the emerging concept is that a specific DDX may interlink more steps of RNA metabolism by performing different activities and recruiting functionally different protein partners to the same selected target RNAs. In fact, DEAD-BOX proteins are thought to function as adaptors or bridging factors in the many steps of RNA metabolism in which they are involved. This notion agrees with the observation that highly specific regions of individual proteins that lie in the N- and C-terminal regions flanking the conserved core, are also the ones that interact with other factors. These interactions could determine their activity in different cell types and/or in response to different signals [134]. Interacting cofactors can stimulate the ATPase and helicase activities, as well as modulate the affinity of the helicase toward specific substrates. In some cases, this interaction has also been shown to be inhibitory [144, 163].

The mechanism by which DDXs recognise their targets is mostly unknown. However, a few studies have shown that they can be selectively recruited through direct binding to specific RNA secondary structures [164, 165]. Intriguingly, DDXs have been found to be recruited to a large number of target RNAs but only appear to affect a subset of them [166]. A major challenge in the field is to identify the substrates that are selectively regulated by each RNA helicase at each step of gene expression. Biochemical approaches including UV crosslinking-based techniques like CLIP (UV-crosslinking and Immunoprecipitation) can improve our understanding of the selectivity of DDX RNA helicases.

Given the widespread activity of DEAD box helicases in regulating gene expression at different levels, and the numerous studies reporting their involvement in cellular growth, it is not surprising that alterations in their expression or function can potentially affect normal cellular homeostasis and contribute to cancer development and progression. DEAD box proteins are mutated or deregulated in multiple cancers (Figure 1.7B).

These proteins have been suggested to exert both oncogenic and tumour-suppressive roles, in a context- and cancer- dependent fashion and through different mechanisms that affect a wide range of molecular events, from transcription to translation. This evidence makes DDX enzymes potential candidates for therapeutic development as well as disease biomarkers [154]. Inhibitors that affect RNA binding and helicase activity of DEAD-box proteins, have already been designed, tested and shown to possess anticancer potential [167].

A major strategy has involved the design of small molecule inhibitors targeting the NTPase activity (e.g. nucleoside analogues), small molecules affecting RNA binding and helicase activity, as well as compounds able to interfere with specific protein- protein interactions involving the helicases or specifically directed against post-transaltionally modified DDXs [167, 168].

Cancer cells are often dependent on certain cellular pathways more than normal cells, which makes cancer vulnerable to the targeting of factors that act in specific pathways. This is the case of the translation machinery (as already discussed in Section 1.1.3), to which cancer cells are addicted for both elevated levels of proteins synthesis and expression of specific oncogenes [98]. Among the most prominent examples of targeting the translational machinery, are the inhibitors designed against DDX2 (e.g eIF4A). Many of these compounds have shown anti-tumorigenic ability [169-171] and have now reached the preclinical study stage.

Another interesting case is the one of DDX5, whose phosphorylation at tyrosine residues was shown to mediate oncogenic phenotypes in cancer cells [172-175]. This led to the development of a phospho-DDX5 specific compound whose action is therefore minimally toxic in normal cells and directed to cancer cells were phosphorylation of DDX5 is present and indispensable for overgrowth [176-179] .

1.4 DDX54: a transcriptional and splicing regulator

The DEAD box RNA helicase DDX54 is one of the least characterised members of the of DDX protein family. In humans, DDX54 is mainly found in the nervous system, gastrointestinal tract organs, female breast and placenta (The Human Protein Atlas-https://www.proteinatlas.org/ENSG00000123064-DDX54/tissue). In most tissues DDX54 has both nuclear and nucleolar sub-localization.

In addition to the common and conserved domains of the DEAD box protein family, DDX54 contains two bipartite nuclear localization signals and three nuclear receptor boxes (L*XX*LL motifs). DDX54 was first identified as an interactor of Estrogen Receptor α (ERα). In fact, DDX54 interacts *in vitro* with ERα and represses estrogen-bound ERα transcriptional activity through a small region (amino acids 589 – 631) that has homology to functional domains of other corepressor proteins (NCoR2/SMRTe) (Figure 1.8). The co-repressive activity of DDX54 was also found to target other nuclear hormone receptors such as progesterone receptor (PR), glucocorticoid receptor (GR) and retinoic acid receptor (RAR) [180].

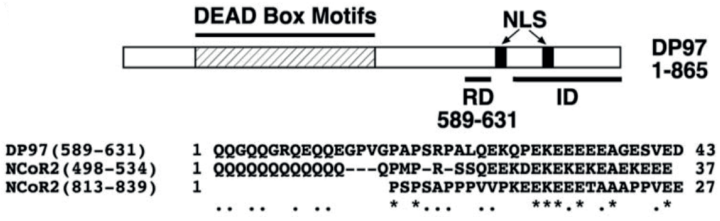


Figure 1.8. DDX54/DP97 structural features and homology with nuclear co-repressors.
Protein BLAST analysis of aa 413– 656 region of DP97 (yeast homolog of DDX54) showed significant sequence similarity to two repressor domains (RD) in co-repressor NCoR2. Asterisks (*) indicate the same amino acid in DP97 and both RDs of NCoR2; bold dots (•) indicate same amino acid in DP97 and one of the RDs in NCoR2. (ID, receptor interaction domain) [180].

Conversely, another study showed that DDX54 acts as a co-activator of the constitutive androstane receptor (CAR). Direct interaction of DDX54 and CAR selectively enhances expression of CAR target genes, which mainly encode enzymes involved in drug metabolism. Suppression of CAR target genes was observed in DDX54-depleted human hepatoma cells (HepG2), whereas expression of the same genes was enhanced upon over-expression of DDX54 in the same cell line [181].

The presence of DDX54 in the nucleolus of cells suggests regulatory roles in rRNA processing and ribosome biosynthesis. In fact, a screen for pre-rRNA processing factors identified DDX54 as a determinant factor in ribosome biogenesis [182]. Consistent with previously reported nucleolar localization and function in pre-rRNAs processing, DDX54 was shown to be essential for the early stages of 18S rRNA maturation [183].

More recently, DDX54 was found to regulate alternative splicing of pre-mRNAs of genes up-regulated upon DNA damage. Specifically, upon exposure to ionising radiation, DDX54 is relocated from the nucleolus to the nucleoplasm where it interacts with spliceosome-associated proteins, including U2AF1. DDX54 binds to introns of DNA damage responsive newly transcribed pre-mRNAs and regulates effective splicing of these transcripts. Indeed, DDX54 was required for efficient post-transcriptional gene expression regulation upon DNA damage [183]

1.5 DDX3X: a multifunctional DEAD box RNA helicase

1.5.1 Conservation and structure of the Ded1/DDX3 subfamily of DEAD-box RNA helicases

Ded1/DDX3 is one of the most widely studied DEAD-box RNA helicase subfamilies. Ded1/DDX3 proteins are conserved in eukaryotic organisms from yeast to human and regulate several different aspects of RNA metabolism [146, 184].

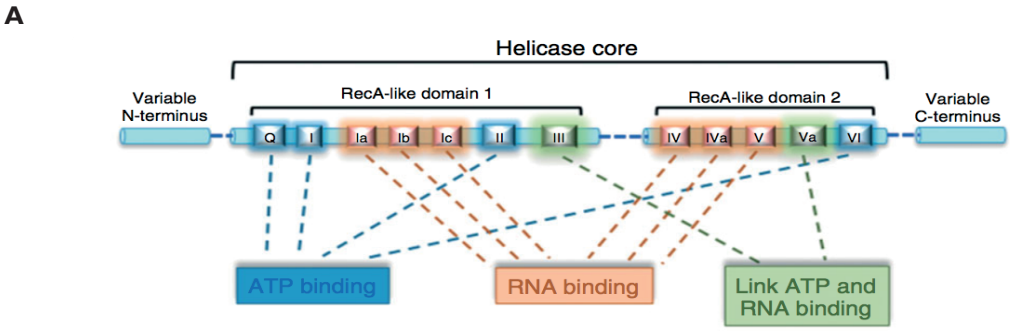
The mouse genome encodes three homologues, namely Ddx3x, Ddx3y and the autosomic homolog Pl10 [185]. The human genome encodes two DDX3 homologues: DDX3X and DDX3Y. DDX3X (or DBX) is an X chromosome-linked gene located in the Xp11.3-11.23 region that escapes X-inactivation and encodes a 662- or 661-amino acid polypeptide, depending on alternative splicing. DDX3Y (or DBY) is located in the *AZoospermia Factor a* (*AZFa*) region of the Y-chromosome [186]. The intron–exon structure of the human *Ded1/DDX3* subfamily of genes is largely conserved throughout higher eukaryotes, suggesting a common ancestor [186, 187].

The mouse and human DDX3 homologues possess non-redundant functions, as suggested from the different expression pattern of these proteins. In humans, *DDX3X* is ubiquitously transcribed in all tissues whereas *DDX3Y* is only expressed in male germ cells [188]. In contrast, amongst the three mouse homologs, *Ddx3x* and *Ddx3y* are expressed in all analyzed tissues whereas the autosomic homolog *Pl10*, located on chromosome 1, is only expressed in testis [189].

The characteristic motifs of the helicase core are highly conserved from yeast to human among the members of the Ded1p/DDX3 subfamily and are divided into the two recA-like helicase core domains. Among others, human homologues contain an additional 10 amino acid insertion (aa 250–259) that is not found in other eukaryotes (*Caenorhabditis elegans*, *Drosophila melanogaster* and *Saccharomyces cerevisiae*) and is important for ATP binding, ATPase activity, nucleic acid binding and unwinding. Indeed, this insertion is located between motifs I (ATP-binding) and Ia (RNA-binding) and it is thought to play a role in positioning the 3′ end of the RNA substrate close to a region of positively charged groups (Figure 1.9 A-C) [146, 190].

As shown in Figure 1.9D, in addition to the recA-like core domains, other regions of sequence conservation are also found within members of this

family. Regions immediately adjacent to the helicase core are conserved between DDX3, Ded1, and Vasa/DDX4. The N-terminal extension (NTE; residues 132–168) is predicted to form a short-helix from residues 145 to 151, whereas the C-terminal extension (CTE; residues 582– 607) has no predicted structure but is highly positively charged. In the latest DDX3 crystal structure, Floor and colleagues have included the N- and C-terminal conserved extension of the helicase core domain (NTE and CTE). Interestingly, they showed that the DDX3 crystal structure, including NTE and CTE portions, has a unique closed conformation forming an interdomain interface which involves the helicase core, and clashes with the RNA-binding surface. This conformation of the DDX3 monomer bound to AMP, suggests that the protein is normally in an auto-inhibitory state and refractory to RNA binding (Figure 1.9E). In fact, mutation of residues in this interface increased the duplex unwinding rate *in vitro*. This suggested that the RNA-dependent rearrangement of the N- and C-terminal flanking tails is needed for full DDX3 ATPase activity [191].



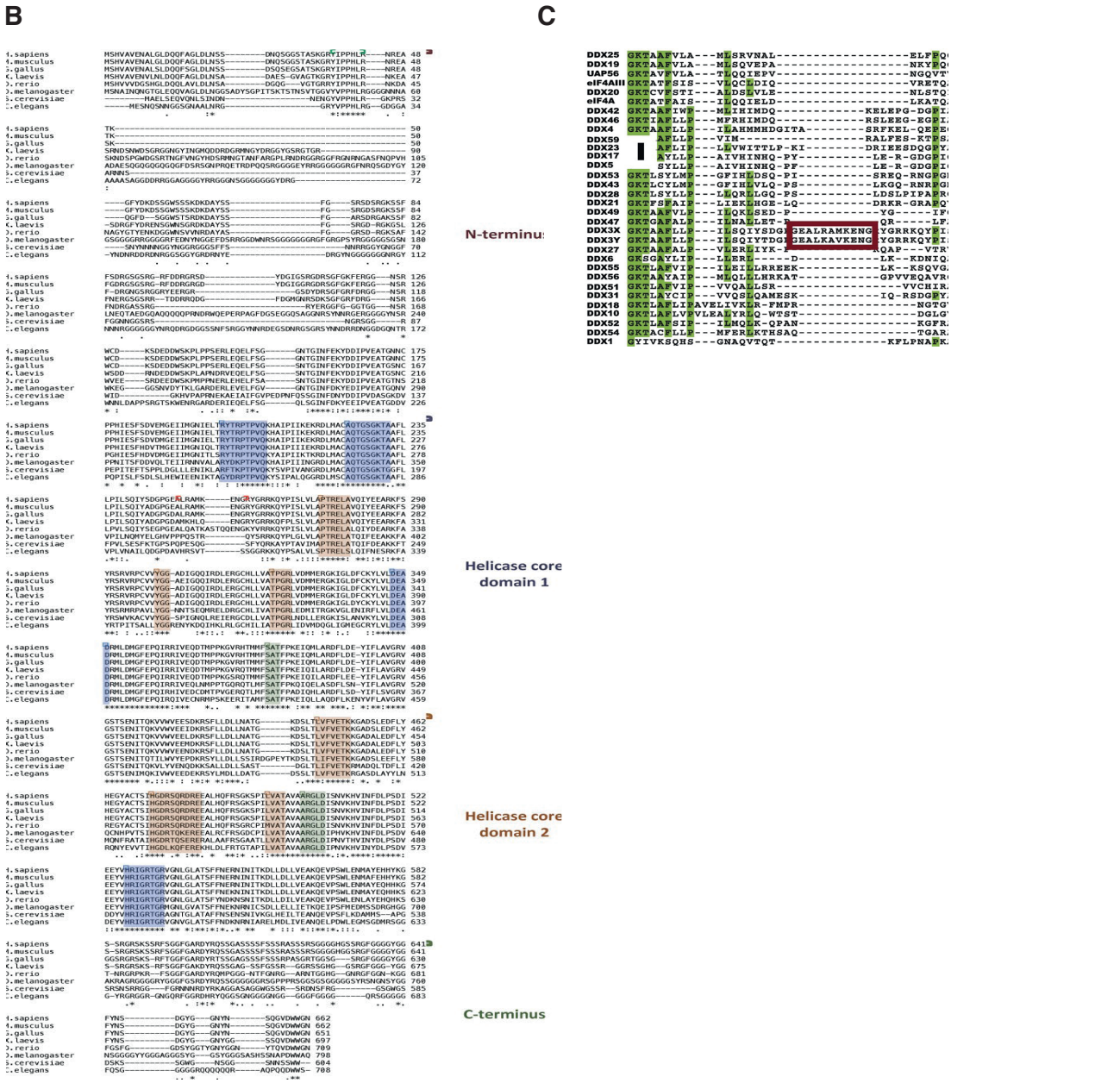
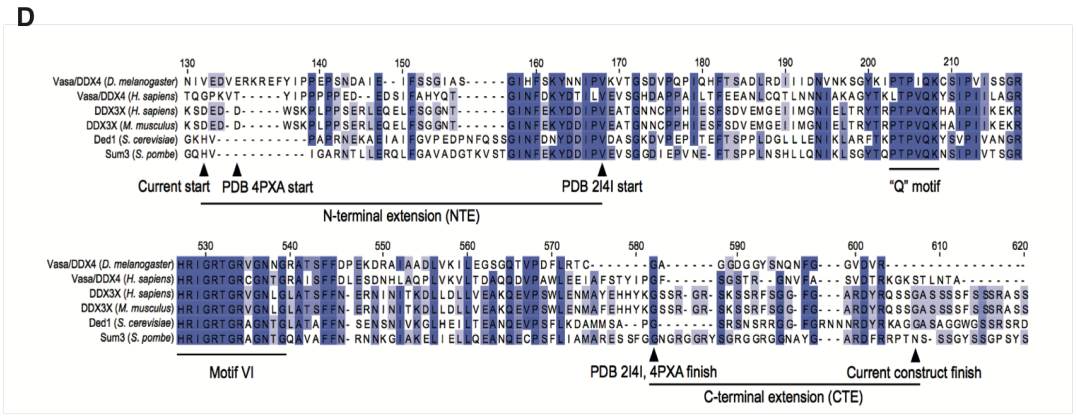


Figure 1.9. DDX3 sequence conservation and functional domains.
(A) Schematic representation of DDX3 domains. The two recA-like domains (domain 1 and domain 2) contain the SF2 conserved motifs involved in ATP binding (blue), RNA binding (brown), and linking of ATP and RNA binding functions (green). The variable N- and the C-terminus regions are also indicated. (B) Human DDX3 sequence (AAC34298.1) was aligned with chicken (NP_001025971.1), frog (An3, NP_001095245.1), zebrafish (CAQ14635.1), fruit fly (belle, AAF54262.1), worm (LAF-1, CCG28150.1), and yeast (Ded1, CAA99419.1). DDX3 orthologue sequences were aligned. The conserved catalytic motifs in the helicase core are indicated with the colour code as in A. Additionally, the highly conserved eIF4E-binding motif (green brackets) and the less conserved 10-amino acid insertion (red brackets), are also indicated [146]. (C) Sequence alignment of 32 different human DEAD-box helicases covering the conserved region between motifs I and Ia (amino acid residues 229–279 in DDX3X). Conserved residues are shown in green. Red box shows the DDX3X/Y-specific 10 amino acid insertion (between residues 250 and 259) [190].



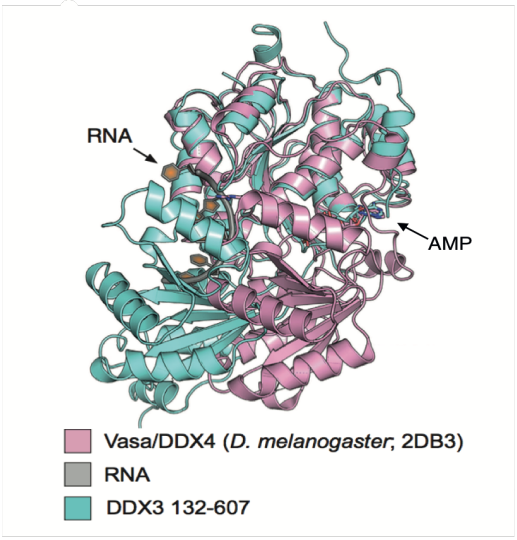


Figure 1.9. DDX3 sequence conservation and functional domains.
(D) Sequence alignments of Vasa/DDX4 (*D. melanogaster*), Vasa/DDX4 (*H. sapiens*), DDX3X (*H. sapiens*), Ddx3x (*M. musculus*), Ded1p (*S. cerevisiae*), Sum3 (*S. pombe*) showing the N- and C-terminal extensions found in the Ded1/DDX3 family and indicating construct boundaries used for the crystal structure. Numbers correspond to human DDX3X.
(E) The 2.2 Å crystal structure includes the conserved core of wild-type DDX3 and the CTE and NTE conserved extensions. The structure of DDX3 (132– 607) bound to AMP (*blue*) is in a partially closed state and clashes with the RNA-binding site shown in the overlapping structure of the DEAD-box protein Vasa bound to RNA (Vasa: *pink*; RNA: *grey;* DDX3: *blue*) [191].

In fact, the authors further showed that removal of the CTE region strongly suppresses duplex unwinding *in vitro.* Since Ded1 oligomerization *in vitro* was shown to depend on the C-terminal tail of Ded1 [192], the authors speculated that removal of CTE negatively affects hDDX3 ability to interact with other proteins. These interactors are likely required to open DDX3 conformation and allow RNA binding.

The N-terminal tail of DDX3 contains a CRM1-dependent nuclear export sequence [193] and an eIF4E-binding motif [194], whereas the C-terminal tail contains an RDYR motif, an invariant WW dipeptide motif conserved sequences and an RS-like domain, all of mostly unknown function [186].

1.5.2 Roles of DDX3X in RNA metabolism

Functional studies on human DDX3 proteins have almost exclusively focused on DDX3X implicating this protein in multiple processes of cellular RNA metabolism. Despite the predominantly cytoplasmic localization of DDX3X, which is consistent with a translational regulation function, there is also evidence for nuclear functions of the helicase, including roles in transcription, pre-mRNA splicing, mRNP assembly, as well as mRNA export [146, 186]. In the following sections, I summarise the current knowledge concerning the roles of DDX3X in post-transcriptional processes.

(i) Pre-mRNA splicing. The role of DDX3X in pre-mRNA splicing was indicated by proteomic studies that identified DDX3X in purified human spliceosomes [195]. Moreover, DDX3X was found to associate with spliced mRNPs in a exon-junction complex (EJC)-dependent manner [196]. It has been hypothesized that DDX3X might play a role in pre-mRNA processing at splicing junctions but evidence for the importance of DDX3X in splicing regulation is still lacking, since no altered splicing events have been reported as a consequence of DDX3X depletion.

(ii) Nuclear RNA export. DDX3X has been associated with mRNA nuclear export involving RNA export factors NFX1 and CRM1, although there is no conclusive evidence about DDX3X's role in this process. The RS-like domain in the C-terminal region of DDX3X is involved in the interaction of DDX3X with the nuclear export factor NFX1 [197]. This interaction persists during shuttling between the nucleus and the cytoplasm since DDX3X was found associated with NXF1 in both of these cellular compartments. However, since depletion of DDX3X does not affect the distribution of NXF1-dependent mRNAs, the relevance of DDX3X's role in mRNA export is uncertain.

DDX3X has also been implicated in CRM1-dependent export in the context of viral mRNA export. However, data involving DDX3X-dependent export of HIV1 mRNAs are so far controversial. In fact, DDX3X was initially reported to be required for RNA export through interaction with both CRM-1 and the HIV1 protein Rev, which recognizes the Rev responsive element (RRE) on HIV1 intron containing transcripts. But this result was questioned by a later study showing that DDX3X knock down does not affect Rev dependent export of HIV1 mRNAs [193, 198].

(iii) Translation. DDX3X has been shown to interact with members of the translation initiation complex eIF4F [199]. In particular, it was shown to directly interact with eIF4G, eIF4E and PABP in co-immunoprecipitation experiments, as well as in *in vitro* pull-down assays [200, 201].

Interestingly, both DDX3X-mediated repressive and activating translational activities have been reported. Interaction of DDX3X with eIF4A was required for repression of cap-dependent translation of reporter constructs [201]. Conversely, interaction between DDX3X and eIF4G promotes translation of a subset of viral RNAs that harbour a stem-loop secondary structure in their 5' UTR [198].

Moreover, evidence for DDX3X activity in translational regulation of selected cellular mRNAs has been reported. Notably, human *cyclin E1* mRNA translation is specifically dependent upon DDX3X binding to the

5'-UTR of the *cyclin E1* transcript. This was demonstrated in both polysome-associated mRNA analyses as well as in *in vitro* assays [202]. Depletion of DDX3X from HeLa cells decreased the level of cyclin E1 protein but not the level of other cyclins [202].

The notion of selective translational activity of DDX3X was challenged by reports showing that DDX3X catalytic mutations actually impact global translation. DDX3X catalytic mutations (occurring in medulloblastomas) do not affect the ability of the protein to bind its target mRNAs but cause impairment of translation of DDX3X target and non-target transcripts, as shown in ribosome profiling studies [203]. Interestingly, this global translational impairment was associated with abnormal formation of stress granules in which translationally repressed mRNAs were sequestered through DDX3X binding [203].

All of the above lines of evidence, implicating DDX3X in translational regulation, leave many open questions: does DDX3X stimulate rather than repress translational processes? Does DDX3X act preferentially as a general translation factor rather than being targeted to specific structures or sequences in specific transcripts? Is DDX3X specifically required for translation initiation or is it involved in elongation as well?
These are non-mutually exclusive possibilities and at present none can be ruled out. It is therefore legitimate to speculate that the molecular function of DDX3X in translation could respond to the cellular context and act differently within different signalling pathways.

(iv) Cytoplasmic mRNA granules. DDX3X is found within stress granules (SGs) formed upon environmental stress and DDX3X depletion reduces assembly of SGs [203]. DDX3X-dependent stress granule aggregation does not require the catalytic domain of DDX3X but instead requires the N-terminal domain involved in interaction with eIF4E. Based on this observation, DDX3X is believed to be located to SGs through interaction

with members of the translation initiation machinery (eIF4E and PABP1) rather than acting as a SG nucleating factor itself [200].

1.5.3 Multifunctional biological activity of human DDX3 proteins: relevance of DDX3X to cancer biology

As described earlier, the human genome encodes two DDX3 genes, one on each of the sex chromosomes. DDX3Y is located in the *AZoospermia Factor a* (*AZFa*) region of the Y-chromosome and appears to be critical in spermatogenesis and required for early stages of testis development in human foetal germ cells. In fact, deletions in the *AZFa* region of the genome cause a condition called Sertoli Cell-Only Syndrome (SCOS) characterized by a complete loss of germ cells in the testis that leads to male infertility [204, 205]. Deregulated expression of DDX3Y has also been linked to testicular tumours [205].

Most research on the biological functions of the DDX3 human subfamily has concentrated on the DDX3X homologue. As highlighted above, DDX3X regulates RNA metabolism at many levels. Moreover, DDX3X plays important roles in many cellular processes in cancer cells including hypoxia, apoptosis, cell cycle regulation and epithelial to mesenchymal transition [206]. Altered expression of DDX3X, or mutations at the *DDX3X* locus, are associated with several different tumours. These findings have led DDX3X to be described as both a tumour suppressor and an oncogene, making it complex to consider DDX3X as a target for cancer therapy [207].

The role of DDX3X in cancer biology has been studied in numerous different cancer models. The mechanisms through which DDX3X regulates cell growth, cell cycle progression and cell motility pathways have only been partially uncovered since most of the available evidence is limited to correlations between expression of DDX3X, or mutations in the *DDX3X* locus, and different types of cancer [206].

In the next two sections, I summarise the available knowledge concerning the tumour suppressive and oncogenic role of DDX3X in different cancers.

(i) DDX3X as a tumour suppressor. Loss of function mutations due to missense and nonsense substitutions in the *DDX3X* gene are associated with head and neck squamous cell carcinomas [208]. Furthermore, low expression of DDX3X at both the RNA and protein level, correlates with poor prognosis and metastasis formation in colorectal cancer, human papilloma virus (HPV)-associated lung cancer and Hepatitis B virus (HBV)-positive hepatocellular carcinoma (HCC) patient samples, compared to the corresponding non-tumour tissues [209-211].

In *in vitro* studies, forced expression of DDX3X suppresses colony formation ability of HCC, cervical carcinoma and colon cancer cells through up regulation of the p21 cell cycle inhibitor [212].

(ii) DDX3X as an oncogene. DDX3X has mainly been shown to act as an oncogene in many cancer cell lines. This activity has been related to different biological processes, especially cell cycle progression and cell motility.

Gain of function mutations have been identified in medulloblastoma subtypes. In fact, *DDX3X* is mutated in 50% and 11% of pediatric WNT- and SHH-subtypes, respectively [213] and in 54% of adult SHH-subtype [214]. All medulloblastomas- associated *DDX3X* mutations reported so far, are nonsynonymous single nucleotide substitutions localized in the helicase core and predicted to alter the normal function of the protein rather than inactivating it [203]. For example, missense mutations localized across the RNA binding domain and the DEAD-BOX domain of *DDX3X,* were found in WNT-subgroup medulloblastomas and predicted to alter the specificity or affinity of the protein for its target RNAs [215].

High levels of DDX3X protein expression correlate with shorter overall survival and metastasis formation in patient samples from different histological subtypes of lung cancer and in colorectal cancer patients [216, 217].

In *in vitro* studies, knockdown of DDX3X significantly reduces colony formation and proliferation of lung cancer cells, as well as causing cell cycle arrest of these cells in G1. The same phenotypic consequences are observed upon DDX3X knockdown in colorectal cancer cell lines [216, 217]. Knockdown of DDX3X in Ewing sarcoma cells results in impaired oncogenic activity as shown by slowed proliferation. In *in vivo* studies, where control and DDX3X knockdown Ewing sarcoma's cells were implanted subcutaneously in recipient mice, tumour volumes were significantly smaller in knockdown cells implants compared to control cells implants [218].

Thus, data concerning DDX3X function in cancer are conflicting. The multifunctional biological activity of DDX3X may explain why this protein could exert very different functions in different cell types and/or in different conditions.

1.6 Role of DDX54 and DDX3X in breast cancer and leukaemia

The following sections are focused on DDX54 and DDX3X function in breast cancer and leukaemia cells, which are the tumour systems relevant to the experimental work described in this thesis.

1.6.1 Role in breast cancer

Breast carcinoma is a very heterogeneous disease classified on the basis of histopathological and genetic features that affect both the prognosis and the choice of the therapy [219]. Most breast cancers derive from the epithelium of the breast with different subtypes showing variable degrees of invasiveness and metastatic potential.

Expression of certain surface, cytoplasmic or nuclear receptors is one of the discriminants of breast cancer subtypes. Breast cancer cells may or may not express three important receptors: oestrogen receptor α (ERα), progesterone receptor (PR) and human epidermal growth factor 2 (HER2) [219]. Most breast cancer *in vitro* studies involve the use of cell lines of different tissue of origin and properties. The most widely used are (i) MCF7 cells, which are human breast epithelial cancer cells characterized by ERα expression and oestrogen growth-dependence, (ii) MDA-MB-231 cells, which are human breast epithelial cancer cells characterized by high invasiveness, and (iii) MCF10A cells, which are immortalized non-tumorigenic human breast epithelial cells commonly used as a reference cell line.

As described earlier (Section 1.3), DDX54 interacts with ERα in MCF7 cells when the receptor binds oestrogen. Binding of ERα to oestrogen determines DDX54 activity as a transcriptional repressor. The expression of genes activated by oestrogen-bound ERα is enhanced by DDX54 depletion whereas the repression of genes inhibited by oestrogen-bound ERα genes was attenuated. Although DDX54 represses the transcriptional activity of ERα on a variety of promoters, it does not interfere with ERα binding to oestrogen responsive elements on DNA [180]. In MCF7 cells, DDX54 is essential for cell survival upon DNA damage as it ensures efficient splicing of DNA damage responsive transcripts [183].

Over-expression of DDX3X in MCF10A cells results in increased invasive properties [220]. Conversely, knockdown of DDX3X decreases cell proliferation and clonogenic cell capacity in MCF-7 and MDA-MB-231 cells. In addition, decreased metastasis formation was observed in a preclinical breast cancer xenograft mouse model in which DDX3X knockdown MDA-MB-231 cells were transplanted [221].

Hypoxia is a major characteristic of solid tumours, which greatly impacts cellular and tumour metabolism. Gene expression changes underlying tumour adaptation to hypoxic conditions are mainly due to the activity of hypoxia inducible factors (HIFs). DDX3X expression contributes to the survival of breast cancer cells when they are subjected to hypoxic conditions. Indeed, the *DDX3X* promoter is a direct target of HIF-1α dependent transcriptional activation [222]. Moreover, DDX3X is overexpressed in breast cancer patient samples, and this correlates with over-expression of HIF-1α and its downstream genes [223].

The role of DDX3X in breast cancer cells during DNA damage has also been investigated. In p53 wild type breast cancer cells, such as MCF7, DDX3X associates with p53 and stabilises it during the DNA damage response [224]. This interaction is necessary for induction of p21 expression and the downstream cascade that leads to apoptosis. Interestingly, in the same conditions, but in p53 mutant cells, such as MDA-MB-231 cells, DDX3X inhibits apoptosis and is essential for survival [224]. This links the action of DDX3X during DNA damage response to the p53 expression status of the cells.

Another study identified high expression levels of cytoplasmic DDX3X in distant breast cancer metastases, especially in brain lesions. This correlated with a worse overall patient survival [225].

The majority of DDX3X is found in the cytoplasm, but a subset of tumours has distinct nuclear DDX3X localization. In a very recent study, DDX3X was detected in the nucleus in breast cancer patient samples and was particularly high in the nucleolus. Nuclear DDX3X correlated with worse overall patient survival. Moreover, nuclear DDX3X localization is due to decreased CRM1-mediated export which requires interaction of CRM1 with the nuclear export signal (NES) located in the DDX3X N-terminal domain [226].

1.6.2 Role in leukaemia

Leukaemia is a progressive malignant disease that affects the blood, bone marrow, lymphatic system, the haematopoietic and lymphoid tissues. It is characterized by abnormal proliferation of dysfunctional and undifferentiated leukocytes, resulting in an unbalanced production of blood cells. Different subtypes of leukaemia affect different blood cells and can be broadly grouped into four types: acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), and chronic myeloid leukaemia (CML) [227-230]. Whilst no evidence for a role of DDX54 in any type of leukaemia has been reported so far, DDX3X has been found mutated, or involved in chromosomal translocations, in different subtypes of leukaemia [231-234].

Truncating mutations (either nonsense mutations or frameshift indels) in *DDX3X* have been associated with CLL cases and reported to cause loss of the protein expression [231]. Recurrent inactivating mutations (nonsense, frameshift and copy-loss mutations) localized in the helicase core and leading to truncation or loss of DDX3X protein, were found in natural killer/T-cell lymphoma patients [232]. Both the previous cases would account for a tumour suppressor role of DDX3X in these leukaemia subtypes.
Moreover, *DDX3X* is involved in the MLLT10-DDX3X gene fusion associated with paediatric T-ALL. Interestingly, the fusion causes loss of DDX3X's DEAD-box domain [233, 234]. More recently, a CRISPR/Cas9 screen showed a subset of AML cell lines are vulnerable to DDX3X knockout [235].

During my PhD, I focused on the two human DEAD-box helicases DDX54 and DDX3X and aimed to investigate the role of these proteins in cancer cells. The main biological system I chose for my research was the breast cancer cell line MCF7. As mentioned, both DDX54 and DDX3X have documented roles in breast cancer cells, particularly MCF7 cells.

Specifically, the role of DDX54 has only been superficially studied in this cell line and, although evidence concerning DDX3X's role in breast cancer indicates an oncogenic function, a deep understanding of the molecular and mechanistic processes underlying the described phenotypes is currently absent.

In the final stages of my research I investigated the role of DDX54 and DDX3X in AML mouse and human cell lines. A role for DDX54 and DDX3X in AML is currently not known.

Chapter 2 - Materials and Methods

2.1 General materials

2.1.1 Standard solutions

Table 2.1 shows the composition of bacterial media and stock solutions used.

Name	Composition per litre
LB	5 g NaCl, 5g yeast extract, 10 g bacto-tryptone
TE	10 mM Tris-HCl, 1 mM EDTA, pH 8.0
TBE	90 Tris-borate, 2 mM EDTA, pH 6.5
TBS	150 mM NaCl, 20 mM Tris-HCl pH 7.6
PBS	150 mM NaCl, 2.5 KCl, 10 Mm Na_2HPO_4, 2 mM K_2HPO_4
SSC 20X	0.3 M sodium citrate pH 7,3 M NaCl

Table 2.1. Stock solutions used.

2.1.2 Vectors

The following plasmids were used in various applications:
pPB-CAG-hKLF4-pA-pgk-Hygro and pGG131 were a kind gift from Dr Meng Amy Li and Dr Ge Guo (Cambridge Stem Cell Institute). U6-sgRNA-EFS-GFP plasmid was a kind gift of Jungwei Shi (Cold Spring Harbor Laboratory), the pKLV2-U6gRNA5 (BbsI)-PKGpuro2ABFP-W was a gift from Konstantinos Tzelepis (Wellcome Trust Sanger Institute). pCMV-VSVg and psPAX2 were purchased from Addgene.

2.2 Bacterial strains and transformation

2.2.1 *Escherichia coli* strain.

DH5α: competent strain for maximized transformation efficiency.

2.2.2 Growth of bacterial cultures

Plated cultures were grown at 37° C overnight on LB plates (LB with 20g bacto-agar per litre). Bacterial suspension cultures were grown in LB at 37° C overnight. To allow the selection of bacteria transformed with plasmid DNA expressing the gene of interest, ampicillin (100 µg/ml) was included in the media.

2.2.3 Transformation of *E. coli* by heat shock.

For the transformation of *E. coli*, 0.1 µg of DNA was added to 100 µl of competent bacteria and incubated on ice for 30 minutes. Samples were placed for 45 seconds at 42° C, returned to ice for 1 minute following which 1 ml of LB was added. This was incubated at 37° C for 1 hour. Bacteria were then plated immediately.

2.3 Preparation of DNA

The QIAGEN Maxi-prep and Mini-prep kits were used according to the manufacturer's instructions for large and small scale plasmid DNA preparation, respectively.

2.4 Cloning of target guide sequences (sgRNA) for CRISPR/Cas9.

The mouse sgRNAs (table 2.2) were designed by Isaia Barbieri (Kouzarides Lab-University of Cambridge) and Jungwei Shi (Vakoc Lab-Cold Spring Harbor Laboratory); the human sgRNAs (table 2.2) were designed by Konstantinos Tzelepis (Wellcome Trust Sanger Institute), using the CRISPR design tool at http://crispr.mit.edu/. sgRNAs were cloned into the plenti-U6-sgRNA-EFS-GFP vector (mouse sgRNAs) or pKLV2-U6gRNA5-(BbsI)-PKGpuro2ABFP-W vector (human sgRNAs) following standard target guide sequence cloning protocol.

Briefly, the vector was digested with BsmBI (GFP vector) or BbsI (BFP vector) for 30 minutes at 37° C, with the addition of FastAP alkaline phosphatase for 5' ends dephosphorylation. Digested vectors were run on a 1% agarose gel and purified by excising the relevant gel band and purification using QIAquick Gel Extraction kit following the manufacturer's instructions.

sgRNA oligo pairs (Table 2.2) were phosphorylated and annealed in the following reaction:

- 1 µl 100 µM oligo1
- 1 µl 100 µM oligo2
- 1 µl 10X T4 Ligation Buffer (NEB)
- 6.5 µl ddH2O
- 0.5 µl T4 PNK (10,000 units/ml NEB)

Reaction samples were run with the following parameters:

37° C for 30 minutes

95° C for 5 minutes – ramp down to 25°C at 5° C /min.

Annealed oligos were used at a 1:200 dilution for ligation reactions with digested/dephosphorylated vector, using the Quick Ligase kit (NEB) following manufacturer's instructions.

Ligated constructs were transformed into DH5α bacteria for plasmid amplification and extraction (Section 2.3).

sgRNA target gene	sgRNA oligos sequences (5'→3')
Ddx54 #1	Forward: caccGTCAGCCGCTCAAACATCGGG Reverse: aaacCCCGATGTTTGAGCGGCTGAC
Ddx54 #2	Forward: caccGCACATGCACCAGACGCCCAG Reverse: aaacCTGGGCGTCTGGTGCATGTGC
Ddx3x #1	Forward: caccGCGCAAGGACGGACTCTAGAT Reverse: aaacATCTAGAGTCCGTCCTTGCGC
Ddx3x #2	Forward: caccGCAGATTCGAGACTTAGAACG Reverse aaacCGTTCTAAGTCTCGAATCTGC

Rpa3	Forward: caccGGCTGGCGTTGACGCGCGCTT Reverse aaacAAGCGCGCGTCAACGCCAGCC
Rosa26	Forward: caccGGAAGATGGGCGGGAGTCTTC Reverse: aaacGAAGACTCCCGCCCATCTTCC
DDX3X Exon 2	Forward: caccGATCTGAAGAGTTCAGGTCT Reverse: aaacAGACCTGAACTCTTCAGATC
DDX3X Exon 3	Forward: caccGCTCGGTTCCTTAAATGAGG Reverse aaacCCTCATTTAAGGAACCGAGC
DDX3X N-terminal domain #1	Forward: caccGGTGGACGGAGTGATTACGA Reverse: aaacTCGTAATCACTCCGTCCACC
DDX3X N-terminal domain #2	Forward: caccGGGAGTGATTACGATGGCAT Reverse aaacATGCCATCGTAATCACTCCC
DDX3X ATP binding loop #1	Forward: caccGTGAGGGAACTCTTTTCTGG Reverse aaacCCAGAAAAGAGTTCCCTCAC
DDX3X ATP binding loop #2	Forward: caccGCTCCCACCAAGTGAACGCT Reverse: aaacAGCGTTCACTTGGTGGGAGC
DDX3X ATP binding and hydrolysis site #1	Forward: caccGATGCTTTTGCACTGGAGTT Reverse aaacAACTCCAGTGCAAAAGCATC
DDX3X ATP binding and hydrolysis site #2	Forward: caccGGAATAGCATGCTTTTGCAC Reverse aaacGTGCAAAAGCATGCTATTCC
DDX3X RNA binding domain (portion A) #1	Forward: caccGTTGGCAGTACAGATCTACG Reverse: aaacCGTAGATCTGTACTGCCAAC
DDX3X RNA binding domain (portion A) #2	Forward: caccGCACCACCATAAACCACGCA Reverse: aaacTGCGTGGTTTATGGTGGTGC
DDX3X RNA binding domain (portion B) #1	Forward: caccGAAGGGTGCAGATTCTCTGG Reverse: aaacCCAGAGAATCTGCACCCTTC
DDX3X RNA binding domain (portion B) #2	Forward: caccGCCATGGAGACCGTTCTCAG Reverse: aaacCTGAGAACGGTCTCCATGGC
DDX3X RS-like domain #1	Forward: caccGACTACAAGGGTAGCAGTCG Reverse: aaacCGACTGCTACCCTTGTAGTC
DDX3X RS-like domain #2	Forward: caccGGAGACTACCGACAAAGTAG Reverse: aaacCTACTTTGTCGGTAGTCTCC
AURKB #1	Forward: caccGAAAATAGTTGTAGAGACGC Reverse aaacGCGTCTCTACAACTATTTTC
AURKB #2	Forward: caccGATGCTCTAATGTACTGCCA Reverse: aaacTGGCAGTACATTAGAGCATC

Table 2.2. sgRNAs sequences and target genes.
The listed sgRNAs were cloned in the relevant vectors to perform CRISPR/Cas9.

DNA oligonucleotides were synthesized by IDT (Integrated DNA technologies). Automated DNA sequencing was performed by Source Bioscience in Cambridge (UK).

2.5 RNA purification and mRNA quantification

2.5.1 RNA purification

Total RNA was purified from cells using the miRNeasy Mini Kit (Qiagen) following manufacturer instructions. Excess of contaminating DNA was removed by On-Column DNase Digestion with the RNase-Free DNase Set (Qiagen), following manufacturer's instructions.

2.5.2 First strand cDNA synthesis

The Superscript III Reverse Transcriptase kit (Invitrogen) was used according to manufacturer's instructions to synthesise first strand cDNA. Briefly, RNA was mixed with 1 µl hexanucleotide primers (100 ng/µl), 1 µl dNTP (10 mM) and 12 µl H_2O. Samples were heated for 5 minutes at 65°C then mixed with 4 µl 5x first strand buffer, 1 µl 0.1 M DTT, 1 µl of RNAse out (a recombinant ribonuclease inhibitor at 40 U/µl) and 1 µl Superscript III RT for 5 minutes at 25°C. cDNA was synthesised by incubating samples at 50°C for 30-60 minutes. Reactions were inactivated for 15 minutes at 70°C.

2.5.3 Quantitative real-time RT-PCR

cDNA purified as described (2.5.2) was used to quantify mRNA levels. Quantitative real-time PCR was performed with the ABI Prism 7300 (Applied Biosystems) sequence detection system using SYBRGreen as the detector. A 10 µl reaction was used on 96 well plates (5 µl of SYBRGreen master-mix, 2 µl of 1 µM forward and reverse primer mix, 3 µl of 1:20 diluted cDNA). Used primers are listed in Table 2.3.

PCR amplification was performed with an initial step of 2 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C.
Relative quantification was obtained calculating ΔΔCt as follows:
Fold change of normalized target gene expression = $2^{-\Delta\Delta Ct}$
Where:
ΔCt1 = Ct (Target gene in treated) – Ct (Ref gene in treated)
ΔCt2 = Ct (Target gene in control) – Ct (Ref gene in control)
ΔΔCt = ΔCt1 (treated) – ΔCt2 (control)
GAPDH expression was used as reference gene. Standard deviation (error bars) and P-values were calculated using the Prism 7 statistical tool.

Primer for cDNA analysis	Sequences (5'→3')
NEAT1 lncRNA isoform ε and β	Forward: GTGGCTGTTGGAGTCGGTAT Reverse: ATTCACTCCCCACCCTCTCT
PVT	Forward: CTTCCTTTGGGTCTCCCTATG Reverse: AAGGCTCCAGGGAGTATGGT
TET2	Forward: GCAGCACACCCTCTCAAGAT Reverse: TCCCTTACTGCTCTTCCTGGA
NCOA1	Forward: TCCTGCCAGCTTCACTTCAG Reverse: CTGATCGCCTGTTCCTGGTT
MYB	Forward: GCAGGTGCTACCAACACAGA Reverse: GGTGTTCTCCCAAACAGGAA
MTA3	Forward: CCCGAGACATCACCTTGTTT Reverse: ACAGGTCCTCCGAGTGGTACT
TFF1	Forward: TTTGGAGCAGAGAGGAGGCAATGG Reverse: TGGTATTAGGATAGAAGCACCAGGG
GREB1	Forward: CACCGGTGTGCACAAGTTAC Reverse: GTGCCTCCGTGTCTAGCTTC
RPS24	Forward: AGAAAGCAACGAAAGGAACGC Reverse: AGCACCTTTACTCCTTCGGC
DDX54	Forward: TGGGACAGTCAGGACAGGAA Reverse: CCCTTCTTCGTCCGAGTCAC
DDX3X	Forward: ACTATGCCTCCAAAGGGTGTCC Reverse: AGAGCCAACTCTTCCTACAGCC

SOX4	Forward: ACTTTGAGAGCATGTCCCTGG Reverse: GTGCAGTAGTCCGGGAACTC
CALR	Forward: CCTGCCGTCTACTTCAAGGAG Reverse: CCGTAGAACTTGCCGGAACT
JUND	Forward: GAAGACCCTCAAGAGTCAGAACA Reverse: TGTTGACGTGGCTGAGGA
PUS1	Forward: TGGTGAGGACATGAGGAAAATG Reverse: GAATGTGAGAGGGAAGGTGG
RPS2	Forward: CCCACACTGTCCCTTGCAA Reverse: AGCAGCTTCTTAGGCACAGG
RPS3	Forward: ATGCTGAAAAGGTGGCCACT Reverse: ATGAACCGCAGCACACCATA
RPS6	Forward: ACGCTCTGGGTGAAGAATGG Reverse: CTCAGTAGCAGGCGGACAC
CCNA2	Forward: CATGGACCTTCACCAGACCT Reverse: GATTTAGTGTCTCTGGTGGGTTG
CDK2	Forward: CATTCCTCTTCCCCTCATCA Reverse: GCCCCCTCTGTGTTAATAAGC
KLF4	Forward: CCATCCTTCCTGCCCGATC Reverse: CGTCTTCCCCTCTTTGGCTT
CCNE1	Forward: CAGATGAAGAAATGGCCAAAA Reverse: TTTGGGTAAACCCGGTCAT
GAPDH	Forward: TGCACCACCAACTGCTTAGC Reverse: GGCATGGACTGTGGTCATGAG

Table 2.3. Primers used in qPCR.

2.6 Protein purification, detection and analysis

2.6.1 SDS polyacrylamide gel electrophoresis (PAGE)

Proteins were separated in 10% acrylamide gels according to size, by the standard Laemmli procedure of one-dimensional gel electrophoresis under denaturing conditions. Protein samples were denatured by boiling in 1X SDS loading buffer (200mM Tris-HCl pH6.8, 20% [v/v] β-mercaptoethanol, 2% [v/v] SDS, 0.1% [w/v] bromophenol blue, 40% [v/v] glycerol) for 5 minutes, and then loaded onto the gel. The gel was separated by electrophoresis in Tris-glycine running buffer (25mM Tris-HCl, 250mM

glycine, 0.1% [v/v] SDS) at 150-200V. Pre-stained high molecular weight protein markers were run alongside the samples (PageRuler from ThermoFisher Scientific).

2.6.2 Western blotting

Proteins were transferred from polyacrylamide gels to nitrocellulose membranes using the mini-trans blot cell (BIO-RAD). Following electrophoresis, gels were transferred in transfer buffer (50mM Tris, 500mM glycine, 20% [v/v] methanol) to a nitrocellulose membrane (Schleicher and Schuell) by applying a constant current (0.4A) for 45 min. After the transfer, the nitrocellulose membrane was stained with Ponceau S to ensure equal transfer and then placed either in milk blocking buffer (TBS with 0.5% [v/v] Tween-20, 5% [w/v] non-fat milk powder) or BSA blocking buffer (TBS with 0.5% [v/v] Tween-20, 5% BSA), for a minimum of 30 minutes at room temperature. The membranes were then probed with primary antibody diluted in blocking buffer, and incubated on a rocking table overnight at 4°C. The next day membranes were washed three times with TBS with 0.5% [v/v] Tween-20 with gentle agitation at room temperature and then probed with secondary antibody (conjugated to horse-radish peroxidase) dilution in blocking buffer at room temperature for 1 hour. Membranes were then washed three times for 10 minutes with TBS with 0.5% [v/v] Tween-20. Detection of protein bands was carried out using the ECL system (Amersham) according to the manufacturer's instructions. The membranes were then exposed to X-ray film (Fuji) or imaged by Chemicdoc (Imaging systems –BIORAD). Antibodies used are listed in Table 2.4.

In some cases membranes were stripped in order to be re-probed with different antibodies. For stripping, membranes were incubated for 1 hour in stripping solution (2% SDS, 62.5 mM Tris-HCl Ph 6.8, 0.8% ß-mercaptoethanol) in gentle shaking. Membranes were then washed twice in TBS with 0.5% [v/v] Tween-20 for 20 minutes in gentle shaking and blocked as previously described.

2.6.3 Preparation of mammalian cell samples for Western blotting

Cells were lysed by the addition of 1X SDS loading buffer. The lysed cells were transferred to an eppendorf tube and sonicated with a VibraCell probe sonicator (Sonics) for 20 seconds at 22% amplitude. The samples were then boiled for 5 minutes. 5-30µl of whole cell extracts were analysed by SDS-PAGE and western blotting.

2.6.4 Antibodies

The following antibodies were used in the listed concentrations for western blot, immunoprecipitation and immunofluorescence assays.

Antibody	Description	Concentration		
		WB	IP (µg/mg of lysate)	IF
DDX54	Rabbit polyclonal-Abcam- ab76947	1:1000	3-5	
DDX54	Mouse monoclonal –Abcam-ab57711	1:1000		1:500
Fibrillarin	Rabbit polyclonal-Nucleolar Marker Abcam-ab5821			1:200
DDX3	Mouse monoclonal -Abcam-ab196032	1:1000	3-5	
β-tubulin	Rabbit polyclonal-Abcam-ab6046	1:1000		
DDX1	Rabbit polyclonal -Abcam - ab70252		3-5	
DDX24	Rabbit polyclonal -Abcam-ab70463		3-5	
GAPDH	Rabbit Polyclonal Abcam ab9483	1:1000		
ERα	Rabbit monoclonal -Abcam -ab32063	1:1000	3-5	
RPS2	Rabbit monoclonal -Abcam ab155961	1:1000		
RPS3	Rabbit monoclonal -Abcam ab181992	1:1000		

RPS6	Rabbit monoclonal -Abcam ab137826	1:1000		
RPS6-(phospho S240 + S244)	Rabbit polyclonal -Abcam-ab53024	1:1000		
JUND	Rabbit polyclonal-Abcam-ab181615	1:1000		
CDK2	Mouse monoclonal –Santa Cruz- sc-6248	1:500		
KLF4	Rabbit polyclonal-Abcam-ab106629	1:1000		
H3 (tri methyl K4)	Rabbit polyclonal –Abcam-ab8580		3-5	
β-tubulin	Rabbit polyclonal - Loading Control Abcam-ab6046	1:500		
GAPDH	Rabbit Polyclonal -Abcam -ab9485	1:2000		
Mouse Control IgG2a	Mouse IgG2a Isotype Control Abcam-ab18413		3-5	
Rabbit IgG, polyclonal - Isotype Control	Rabbit IgG, polyclonal - Isotype Control Abcam-ab27478		3-5	
Alexa594-conjugated IgG.	Rabbit polyclonal, Invitrogen			1:250
Alexa488-conjugated IgG2A	Mouse polyclonal, Invitrogen			1:250

Table 2.4. Antibodies used for immunoprecipitation (IP), western blot (WB) or confocal immunofluorescence (IF)

2.6.5 Co-Immunoprecipitation from mammalian cell extracts (Co-IP).

Cells were washed once in PBS and pelleted by micro-centrifugation (4,000 rpm, 4 min), resuspended in ice-cold IPH lysis buffer (150mM NaCl, 50mM Tris-HCl pH 8.0, 5mM EDTA, 0.5% NP-40) with protease inhibitor cocktail (Complete™, Roche) and incubated on ice for 30 minutes. Extracts were cleared of debris by micro-centrifugation (15 minutes, 13,000rpm, 4°C). Extracts were then incubated with 3µg of antibody, on a rotating wheel at 4°C for a minimum of 2 hours. 20µl protein A or protein G sepharose beads (1:1 mix) were added and the incubation continued for 4-

14 hours. The sepharose beads were pelleted in a cooled micro-centrifuge (4,000 rpm, 1 min, 4°C) and washed 3 times with ice-cold IPH. After the final centrifugation, the beads were resuspended in 1X SDS loading buffer analysed by SDS-PAGE and western blotting.

2.6.6 Fractionation of mammalian cells.

Subcellular fractions were obtained following the REAP fractionation protocol [236]. Briefly cells grown as monolayers in 10 cm diameter dishes were washed in ice-cold phosphate buffer saline (PBS) pH 7.4, scraped from culture dishes on ice using a plastic cell scraper and collected in 1.5 ml micro-centrifuge tubes in 1 mL of ice-cold PBS. After a "pop-spin" centrifugation (for 10 seconds in an Eppendorf table-top micro-centrifuge 13,000 rpm), supernatants were removed from each sample and cell pellets were resuspended in 900 µL of ice-cold 0.1% NP40 (Calbiochem, CA, USA) in PBS and triturated 5 times using a p1000 micropipette (Gilson, WI, USA). 300 µL of the lysate were removed as "whole cell lysate" and 100 µL of 4 × Laemmli sample buffer were added to it, then kept on ice until the sonication step. The remaining (600 µL) material was collected by centrifugation for 10 seconds in 1.5 ml micro-centrifuge tubes and 300 µl of the supernatant were removed as the "cytosolic fraction". 100 µL of 4 × Laemmli sample buffer were added to this fraction and boiled for 1 min. After the remaining supernatant was removed, the pellet was resuspended in 1 ml of ice-cold 0.1% NP40 in PBS and centrifuged as above for 10 seconds and the supernatant was discarded. The pellet (~20 µL) was resuspended with 180 µL of 1 × Laemmli sample buffer and designated as "nuclear fraction". Nuclear fractions and whole cell lysates that contained DNA were sonicated with a VibraCell probe sonicator (Sonics) for 20 seconds at 22% amplitude, and boiled for 1 min. 10 µL, 10 µL and 5 µL of whole cell lysate, cytoplasmic and nuclear fractions, respectively, were loaded and separated by electrophoresis using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.7 Electrophoretic Mobility Shift Assay (EMSA)

2.7.1 Preparation of protein extracts

Cultured adherent cells (MCF7) were washed twice with ice-cold PBS 1X, collected by scraping in 1 ml of ice-cold PBS1X and transferred into a 1.5 ml micro-centrifuge tube. The cell suspension was collected by centrifugation in a micro-centrifuge for 5 min at 700 x g, at 4°C and pellets were lysed in 200 µl of lysis buffer (1% Triton X100, 25Mm Tris-HCl (pH 7.4), 40 mM KCl) supplemented with protease inhibitor cocktail (Complete™, Roche). Lysates were incubated for 20 minutes on ice and then collected by centrifugation at full speed at 4°C. Supernatants were transferred to a new micro-centrifuge tube and kept on ice. Protein content was quantified with a BCA protein assay Kit (ThermoFisher Scientific), following manufacturer's instructions

2.7.2 Preparation of radiolabelled *JUND* RNA probes.

RNA oligonucleotides in Table 2.5 were synthetized by IDT (Integrated DNA technologies). To generate 5' radiolabelled probe, the following T4 polynucleotide kinase reactions were set up:

- 20 pmol of RNA probe
- 1X T4 kinase buffer
- T4 PNK (2 µl – 10,000 units/ml)
- 100pmol ATP, [γ-^{32}P]- 3000Ci/mmol 10mCi/ml

Reactions were incubated for 1h at 37°C with gentle shaking and stopped by adding EDTA, pH8 to a final concentration of 50mM. Labelled probes were purified with RNA Clean & Concentrator™-5, according to manufacturer's instructions. Finally, probes were diluted in RNase-free water (1:10) and quantified by scintillation counter.

2.7.3 Protein extract-RNA probe binding reactions.

30µg of protein extract were combined with 50000cpm/µl labelled RNA probe. Lysis buffer supplemented with 5 U/µl of RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen) was used to dilute the protein extracts to the appropriate amount. In the supershift condition, 1.5 µg of anti-DDX3 antibody were added to the reaction. Binding reactions were carried out at room temperature for 20 minutes. After this time, heparin was added to a concentration of 500 µg/ml and the reaction was allowed to continue for 10 more minutes.

2.7.4 Separation of protein extracts-RNA probes complexes by electrophoresis

Glycerol was added to all samples to a final concentration of 20% (v/v) and samples were run on a 6% TBE gel (Novex- ThermoFisher Scientific) at 60 V for 30 minutes.

Gels were fixed in fixing solution (Methanol 10%, Acetic Acid 20% in H_2O) for 10 minutes at room temperature, then placed into an autoradiography exposure cassette and exposed to film at -80°C for variable exposure times (30 mins to 48h).

RNA probe name	RNA probe sequence (5'→3')
hJUND Wild Type DDX3X Binding Motif	5'-AGCUGGUUCUGCUUGUGUAAAU**CCUCCAGG**GCCUUGACGAAGCCCUCGGCGA-3'
hJUND Mutant DDX3X Binding Motif	5'-AGCUGGUUCUGCUUGUGUAAAU**CGAGCAGG**GCCUUGACGAAGCCCUCGGCGA-3'

Table 2.5. RNA oligo probes used in EMSA.

2.8 Cell lines and mammalian tissue culture.

2.8.1 Cell culture media, reagents and maintenance

Name	Cell type	Other information
MCF7	Human transformed breast epithelial cancer cell line	Estrogen Receptor α +
MCF10A	Human immortalized non-tumorigenic breast epithelial cell line	
OCI AML2/CAS9	Human Acute Myeloid Leukaemia cell line	Stably expressing Cas9.
OCI AML3/CAS9	Human Acute Myeloid Leukaemia cell line	Type A NPM1 gene mutation (NPM1c). Stably expressing Cas9.
RN2c	Mouse acute myeloid leukaemia cell line	MLL-AF9/Nras^{G12D}; Stably expressing Cas9
NIH 3T3/Cas9	Mouse embryonic fibroblasts	Stably expressing Cas9
HEK293T	Human embryonic kidney epithelial cell line	SV40 T-antigen

Table 2.6. Cell lines used.

All tissue culture manipulations were carried out under sterile conditions in a standard laminar flow hood. All cell lines used are listed in Table 2.6. MCF7, NIH-3T3 and 293T cells were cultured in DMEM (Invitrogen), supplemented with 10% FBS and 1% penicillin/streptomycin/glutamine. MCF10A cells were cultured in DMEM/F12 (Invitrogen) supplemented with 5% Horse serum, 20ng/ml EGF, 0.5 mg/ml Hydrocortisone, 100ng/ml cholera toxin, 10 μg/ml Insulin and 1% penicillin/streptomycin. OCI-AML2/Cas9 and OCI-AML3/Cas9 cells were cultured in Alpha-MEM

supplemented with 20% FBS, 1% penicillin/streptomycin/glutamine and 1µg/ml blasticidin.

RN2c cells were cultured in RPMI1640 (Invitrogen) supplemented with 10% FBS and 1% penicillin/streptomycin/glutamine. All cell cultures were stored in incubators at 37°C at 5 %CO_2 (100% humidity) and passaged between 1:3 and 1:5 every 3-4 days to maintain a sub-confluent cell density. Cells were passaged using conventional cell culture techniques. To detach adherent cells, culture medium, and cells were washed in PBS 1X and detached by addition of 1ml/10cm plate Trypsin/EDTA solution (0.025% trypsin and 0.01% EDTA in PBS).

RN2c and NIH-3T3/Cas9 cells were a kind gift from the Vacok lab. These cell lines were derived by retroviral transduction of a murine MLL-AF9/Nras^{G12D} acute myeloid leukemia cell line (RN2) and murine fibroblasts NIH-3T3 with MSCV-hCas9-PGK-Puro, followed by puromycin selection and serial dilution to derive single cell-derived clones. OCI-AML2/Cas9 and OCI-AML3/Cas9 were a kind gift from Konstantinos Tzelepis (Wellcome Trust Sanger Institute). These cell lines were generated by lentiviral transduction using pKLV2-EF1aBsd2ACas9-W, followed by blasticidin selection and serial dilution to derive single cell-derived clones.

2.8.2 Cell proliferation assay

Cells were seeded at 3 x 10^5 in 2 ml of complete medium in three biological replicates. Cells were counted every 2 days after plating using the Countess II cell counter. Proliferation curves and standard deviation were generated using the Prism 7 statistical tool.

2.8.3 Transient transfection of cells using FuGENE 6.

MCF7 and MCF10A cells were plated in a 6-well tissue culture dish in the relevant culture media without antibiotics for 24h reaching 60-80% confluency. Media was replaced with 2ml of Opti-MEM. 6µl of FuGENE™

was mixed with 74µl of Opti-MEM™ (Gibco), 2µg of total DNA was then added to the mix and incubated for 30 minutes at room temperature. Increasing amount of pPB-CAG-hKLF4-pA-pgk-Hygro were transfected (0 µg, 0.5 µg, 1 ug, 2µg) and combined with decreasing amount of pGG131 control plasmid in order to transfect a final amount of 2µg of DNA in all conditions.

The transfection mix was added in droplets to the cells which were left at 37°C, 5%CO_2 for 6-16 hours. After this time, transfection media was replaced by 3 ml of normal media. Cells were harvested 48h after transfection for RNA extraction or processed for cell cycle analysis (Section 2.10.2).

2.8.4 Transient knock down by RNA interference (RNAi).

MCF7 and MCF10A cells were plated in a 6-well tissue culture dish in the relevant culture media without antibiotics for 24h reaching 60-80% confluency. Then, cells were transfected with small interfering RNA oligonucleotides (100 nM final concentration) and Dharmafect I to knock down DDX3X or DDX54 according to manufacturer's instructions (4µl of Dharmafect I were combined with 96µl of Opti-MEM; 10µl of 20µM siRNA were combined with 90µl of Opti-MEM). The siRNA used are listed in Table 2.7.

Target	siRNA name
hDDX3X	J-006874-06
hDDX3X	J-006874-08
hDDX54	J-017128-09
hDDX54	J-017128-10
hDDX54	J-017128-12
hKLF4	J-005089-08
(Non-targeting siRNA) Scramble sequence	D-001810-01

Table 2.7. siRNAs used.

Cells were harvested 72h after transfection for RNA or protein extraction and several downstream assays. All ON-TARGET plus human siRNA and Dharmafect I reagent were purchased from Dharmacon™.

2.8.5 Lentiviral Particles production

For each plasmid to be transfected, HEK293T cells were plated in DMEM + 10% FBS without antibiotics (no penicillin or streptomycin) in a 10 cm tissue culture plate previously coated with poly-D-lysine (Sigma), and cells incubated at 37°C, 5% CO_2 overnight.
A mixture of the 3 transfection plasmids was prepared in a 2:1:0.5 ratio as follows:

- 3 µg pLKO.1 shRNA plasmid/sgRNA containing plasmid
- 1.5 µg psPAX2 packaging plasmid
- 750 ng pVsVg envelope plasmid

and OPTI-MEM medium was added to a final volume of 60 µl.

For each reaction 18 µL FuGENE 6 were mixed with 222 µL OPTI-MEM and 240µl of Fugene mix were added to each DNA mix and incubated for 30 minutes at room temperature. The DNA:FuGENE mix was added dropwise to the cells and medium was replaced with 6 ml of normal fresh DMEM after 16h incubation at 37°C, 5% CO_2. 24h and 48h hours later virus-containing media was harvested from cells, filtered through a 0.45 µm filter to remove the cells and used to infect cells or stored at -80°C.

2.8.6 Generation of stable knock down cell lines through shRNA lentiviral transduction.

Stable DDX3X and DDX54 knock-down MCF7 cells were obtained by transduction of lentiviral particles bearing pLKO.1 plasmids expressing the relevant shRNA and the puromycin selection cassette. The pLKO.1 constructs used are shown in Table 2.8.

Target	pLKO.1-puro ID
hDDX3X	TRCN0000000001
hDDX54	TRCN0000288991
pLKO.1-puro Non-Mammalian shRNA Control Plasmid DNA	SHC002

Table 2.8. shRNAs used.

MCF7 cells were plated in normal DMEM in 10 cm plate to reach 70% confluency the next day. The viral suspension (~ 6 ml) was supplemented with 8 µg/ml human polybrene (Millipore) and added to MCF7 target cells. Cells were left at 37°C, 5% CO_2 overnight. Virus containing media was changed for fresh media 24 hours after infection and to 1µg/ml puromycin-containing media 48h after infection. After three days of selection in puromycin, cells were harvested for RNA or protein extraction.

2.8.7 Generation of DDX54 and DDX3X CRISPR/Cas9 knockout or DDX3X CRISPR/Cas9 domain targeted cell lines.

Lentiviral sgRNA expressing vector (U6-sgRNA-EFS-GFP and pKLV2-U6gRNA5 (BbsI)-PKGpuro2ABFP-W) together with the packaging plasmids PAX2 and VSVg were transfected at a 2:1:0.5 ratio into HEK293T cells for lentiviral production (as described in Section 2.8.5). Supernatant was harvested 48h and 72h after transfection. 1x 10^6 RN2c, OCI-AML2/Cas9 or OCI-AML3/Cas9 cells and viral supernatant were mixed in 2 ml of culture medium supplemented with 8 µg/ml (human) or 4 µg/ml (mouse) polybrene (Millipore), followed by spinfenction (60 min, 900 *g*, 37 °C). NIH-3T3/Cas9 cells were plated in 6 well plates and infected with the same ratio of culture medium:viral supernatant. All cells were further incubated overnight at 37°C. The medium was refreshed on the following day and the transduced cells were cultured further.

For competition assays, a serial dilution of relevant virus in correlation with the GFP+ or BFP+ cell population was used to estimate the viral titre to use. To ensure sub-optimal infection, a 1:3 dilution of the first viral collection (Section 2.8.5) was used under all conditions.

2.9 Immunofluorescence, RNA FISH and confocal microscopy

MCF7 wild type cells or MCF7 cells transfected with either scramble siRNA or DDX54 targeting siRNAs (#9 or 12) at 48h after transfection, were seeded in 12 well culture dishes containing glass coverslips (previously coated with poly-L-lysine and let to attach overnight). The next day cells were washed twice in 1 X PBS then fixed with 4% paraformaldehyde in PBS, pH 7.4 for 15 min at room temperature then washed twice with 1 X PBS. The next steps were diversified to perform either immunofluorescence, RNA FISH or a combination of the two techniques.

2.9.1 Double immunofluorescence

Fixed cells were permeabilized for 10 min at room temperature incubation in 1X PBS containing 0.25% Triton X-100. Cells were washed three times in 1X PBS and then blocked with 1% BSA in PBS-0.1% Tween for 30 min at room temperature. A mixture of anti-DDX54 and anti-Fibrillarin primary antibodies diluted in block solution was added to cells for 1h at room temperature. After primary antibody incubation, cells were washed three times in 1X PBS-0.1% Tween and a mixture of Alexa-Fluor 488 conjugated anti-mouse and Alexa-Fluor 594 conjugated anti-Rabbit secondary antibody diluted in blocking solution was added to cells in the dark for 1h at room temperature. Cells were washed three times with 1X PBS- 0.1% Tween.

2.9.2 RNA Fluorescence In Situ Hybridization (RNA-FISH).

Fixed cells were permeabilised by 1h incubation in 70% ethanol at 4°C, then washed twice in 1X PBS and twice in FISH wash buffer (2X SCC, 10% v/v formammide in Nuclease-free water) for 5 minutes per wash. Next, coverslips were placed cell-side down onto a 100ul drop of hybridization buffer (10% w/v dextran in wash buffer) containing 1:1000 diluted Quasar 570-conjungated NEAT1 lncRNA probes (Stellaris™-LGC) in a humidified chamber. The humidified chambers were sealed and incubated in the dark at 37°C overnight. The next day coverslips were incubated in fresh hybridization buffer in the dark for 30 min at 37°C, to reduce the background, then washed once in wash buffer and twice in 1X PBS.

2.9.3 Immunofluorescence and RNA FISH combined

Fixed cells were permeabilised and blocked as in Section 2.9.2, then washed in 1X PBS and in FISH wash buffer as described in Section 2.9.2 then hybridized overnight with 100µl of hybridization buffer (2.10.2) containing NEAT1 lncRNA probes (1:1000), 0.25% Triton X100 and anti-DDX54 primary antibody. The next day coverslips were washed as described (Section 2.9.2), incubated 1h at room temperature in the dark with anti-mouse Alexa Fluor 488 conjugated secondary antibody solution (Section 2.9.1) and finally washed three times with 1X PBS- 0.1% Tween.
In all cases (Sections 2.9.1-2-3) coverslips were mounted on microscope slides using DAPI-containing mounting medium (Vectashield- Vector Laboratories). Samples were then analysed by confocal microscopy. Confocal laser images were captured with an Olympus Fluoview FV1000 microscope equipped with a 40X oil lens. Image processing was carried out using ImageJ software.

2.10 Flow cytometry

2.10.1 BFP and GFP detection

For detection of GFP or BFP, RN2c, OCI AML2 and OCI AML3 suspension cells were washed and resuspended in 1X PBS at room temperature. For detection of GFP, NIH-3T3 cells were detached by trypsinization, washed and resuspended in 1X PBS at room temperature. Cells were then analysed with a FACSCalibur cytometer (BD) (for GFP detection) or SONY Cell Sorter (for BFP detection). At least 25,000 events per sample were collected. Data were processed with FlowJo software. Standard deviation (error bars) and p-values were calculated using the Prism 7 statistical tool.

2.10.2 Cell cycle assessment

MCF7 cells transfected with either scramble siRNA or DDX54 targeting siRNAs (#9 or 12), or DDX3X targeting siRNA (#6 or #8) were treated with 50ng/ml nocodazole for 16 h (overnight) at 72h after transfection. Cells were detached by trypsinization. OCI-AML2/Cas9 and OCI-AML3/Cas9 cells in which CRISPR/Cas9-mediated excision was performed against the DDX3X RNA binding domain or no target (empty vector) were treated with 50ng/ml nocodazole for 16 h (overnight) at day 8 after infection (day 4 after puromycin selection).

All cells were washed in 1X PBS 10mM EDTA and pellets were resuspended in 100µl 1X PBS 10mM EDTA 1% BSA. 200µl of Propidium Iodide (PI) staining solution (1% NP40, PI 20 µg/ml, RNaseA 0.1 mg/ml in 1X PBS, 10mM EDTA, 1%BSA) were added to resuspended pellets. Samples were kept on ice and measured by a FACSCalibur Cytometer (BD).

At least 25,000 events per sample were collected. Data were processed with FlowJo software. Standard deviation (error bars) and p-values were calculated using the Prism 7 statistical tool.

2.11 RNA-seq

2.11.1 RNA-seq library preparation.

Total RNA was extracted as described (Section 2.5) from MCF7 cells transfected with either scramble siRNA or DDX54 targeting siRNAs (#9 or 12), or DDX3X targeting siRNA (#6 or #8) harvested 72 h after transfection and OCI-AML2/Cas9 and OCI-AML3/Cas9 cells in which CRISPR/Cas9-mediated excision of DDX3X RNA binding domain (A and B portions) or no target (empty vector) harvested at day 8 after infection (day 4 after puromycin selection). All conditions were produced in three independent biological replicates.

Ribosomal RNA was depleted with Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) from Illumina(R), following manufacturer's instructions. RNA seq libraries were produced with the NEXTflex™ RNA-Seq Kit from Bioo Scientific, following manufacturer's instructions. Before multiplexing, excess primer was removed with AMPure XP beads (Beckam Coulter). Before and after multiplexing, libraries were tested for both size and quantity of DNA using the Qubit™ dsDNA HS Assay Kit and High Sensitivity D1000 ScreenTape system following manufacturer's instructions.

2.11.2 RNA-seq bioinformatic analysis

RNA-seq bioinformatic analysis was performed as follows.

(i)For differential gene expression in DDX3X and DDX54 knock down MCF7 cells (Section 2.84):

Trimmed reads were mapped in paired end mode to the h38 human genome using tophat [237] with the following parameters (--no-coverage-search --max-multihits 300 --report-secondary-alignments --read-mismatches 2 --library-type fr-firststrand). Multihits (reads mapping to multi loci) were filtered, along with reads mapping with quality score less

than 20. Reads were counted across gene models taken from the Ensembl v86 gtf gene model list using the *summarizeOverlaps* function from the *GenomicAlignments* package in R [238]. The strand of each read was inverted prior to counting to account for the fact that libraries represent the first strand of synthesised cDNA. Read counts were converted into normalised fragments per kilobase mapped (FPKM) values for quality control plots. Differential expression analysis was conducted on the raw count data using the *DESeq2* package in R [239]. P values were corrected for multiple testing using the Benjamini and Hochberg FDR correction. Significantly changing genes were identified based on a fold change greater than 2-fold (up or down) and an adjusted p value less than 0.05. In addition, significant genes were filtered to remove genes where both the control and mutant samples had an average FPKM score less than 1.

(ii)For differential gene expression in OCI-AML2/Cas9 and OCI-AML3/Cas9 cells upon CRIPSR/Cas9-mediated excision of DDX3X RNA binding domain (Section 2.8.7):

Fastqc files were mapped against the reference genome sequence (GRCh38) using STAR [240] (v2.5.0) with indexes built with the Ensembl [241] reference transcriptome GTF file. The number of reads mapping to each Ensembl annotated gene were counted with STAR and then imported into DESeq2 [239] for differential expression analysis. The DESeq2 workflow was run with default parameters using a one-factor design representing the treatment condition. The resulting p-values were corrected using the Benjamini-Hochberg method as impelemented in DESeq2.

In both cases, Gene Ontology analysis was performed by using default settings of the DAVID tool (DAVID Bioinformatics Resources 6.8 - https://david.ncifcrf.gov - P-values < 0.01).

2.12 Chromatin immunoprecipitation (ChIP)-sequencing.

DNA was cross-linked with 1% (v/v) formaldehyde for 15 minutes at room temperature and cross-linking was stopped by the addition of 0.125M glycine. Cells were then lysed in 1% [v/v] SDS, 10mM EDTA, 50mM Tris-HCL pH8.0, and 1mM protease inhibitors. Cells were sonicated in a QSONICA sonicator (1h cycle 5 secs ON/5 secs OFF) to achieve a mean DNA fragment size of 500bp.

Acceptable fragmentation was checked as follows:

3 times the volume of Elution Buffer (1%SDS, 0.1M $NaHCO_3$) was added to aliquots of non-sonicated and sonicated samples, and samples were boiled at 95°C for 5 minutes

1 volume of phenol:chloroform:isocyamil (25:24:1) was added and samples were vortexed for 20 seconds and collected by centrifugation at RT, 5 min, 13.5 K.

The supernatant was transferred to a fresh tube and DNA was precipitated by addition of 1µl glycogen, 1:10 3M NaAcetate, pH 5.2, adding 2.5 volumes of EtOH 100%, for 1h at -80°C.

Samples were collected by centrifugation for 30 mins, max speed at 4°C, and pellets were washed with 70% EtOH and finally resuspended in H_20.

Samples were incubated with 2µl RNase A with shaking at 37°C for 10 minutes and then run on a 1.5% agarose gel followed by post-staining with EtBr .

An equal volume of protein A and G dynabeads magnetic beads (Thermo Scientific), were equilibrated in modified RIPA buffer (1% [v/v] Triton-X-100, 0.1% deoxycholate, 0.1% SDS, 90mM NaCl, 10mM Tris-HCL pH8.0, and EDTA-free protease inhibitors) and used to pre-clear chromatin under rotation at 4°C for 2 hours prior to immunoprecipitation. After pre-clearing, 100µl of chromatin was mixed with 100µl of lysis buffer and stored as input material. Immunoprecipitation was performed for a minimum of 12 hours at 4°C in modified RIPA buffer with anti-DDX3, anti-DDX54, anti-ERα, anti-H3K4me3 or anti-IgG Rb isotype control

antibodies. All IPs were performed in triplicate. Protein A (for rabbit antibodies) and G (for mouse antibodies) previously equilibrated in modified RIPA buffer, were used to bind the antibody and associated chromatin for a minimum of 2 hours. The beads were then washed twice with wash buffer (0.1% [v/v] SDS, 1% [v/v] Triton-X-100, 2mM EDTA, 20mM Tris pH8.0, 150mM NaCl) and once with final wash buffer (0.1% [v/v] SDS, 1% [v/v] Triton-X-100, 2mM EDTA, 20mM Tris pH8.0, 500mM NaCl). DNA was then eluted with 200µl of elution buffer (1% [v/v] SDS, 100mM $NaHCO_3$) for 15 minutes rotating at room temperature. The input and immunoprecipitated samples were then mixed with 2µl of RNase (DNase free-Roche) for a minimum of 5 hours at 65°C to reverse the crosslinks. DNA purification was performed using the ChIP DNA ZYMO purification kit. Purified DNA was used for preparation of ChIP seq libraries with NEXTflex® ChIP-Seq Library Prep Kit for Illumina® Sequencing from Bio Scientific, following manufacturer instructions. Before multiplexing, excess primers were removed with AMPure XP beads (Beckam Coulter). Libraries were quantified with the Qubit® dsDNA HS Assay and the High Sensitivity D1000 ScreenTape system, following manufacturer's instructions.

2.12.1 ChIP seq bioinformatic analysis.

Trimmed reads were mapped against the hg38 build of the human genome using BWA v0.7.12 [242] [PMID: 19451168] with default parameters. Peaks were called for each replicate IP condition against a reference condition made by combining the data from the Input replicates. Peaks were called using MACS v1.4.2 (Zhang et al. Model-based Analysis of ChIP-Seq (MACS) [243]. Pairwise overlaps between the replicates were taken and the union was used as the final peak set such that a peak exists in the final data set if it is present in at least 2 of the 3 replicates.

2.13 iCLIP (Individual Cross-Linking ImmunoPrecipitation)

iCLIP was performed following the protocol in Huppertz et al 2014 [244].

2.13.1 Cells UVC irradiation and immunoprecipitation, 3' end adapter ligation and 5' end radioactive labelling

MCF7 cells were cultured in 10 cm diameter dishes. When cells reached confluence, media was removed and cells were washed twice with ice-cold PBS 1X. 6ml of ice cold PBS 1X were added, plates were placed on ice and irradiated once at 150 mJ/cm^2, 254 nm. Cells were scraped and harvested into eppendorf tubes and spun at top speed for 10 seconds at 4°C. Pellets were resuspended in CLIP lysis buffer (50 mM Tris–HCl, pH 7.4, 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with protease inhibitor cocktail (Complete™, Roche) and sonicated once at VibraCell probe sonicator (Sonics) for 20 seconds at 22% amplitude. Lysates were treated with 10µl of either high RNase I (Life Technologies) (1:10) or low RNase I (1:500) together with 2µl of Turbo DNase (Life Technologies) for 3 minutes at 37°C shaking at 1,100 rpm and then immediately placed on ice. Samples were then collected by centrifugation for 10 min at 22,000 g at 4°C to clear the lysate. The supernatant was collected and quantified with a BCA assay. Protein A (for rabbit antibodies) or protein G (for mouse antibodies) Dynabeads were previously washed twice in lysis buffer and then resuspendend with 100 µl lysis buffer with the relevant antibody dilution and incubated under rotation at room temperature for 1h. The lysates were added to the antibody/bead mix and incubated overnight under rotation at 4°C.

The next day beads were washed 4 times in High Salt buffer (50 mM Tris–HCl, pH 7.4,1 M NaCl, 1 mM EDTA 1%, Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate) and twice in PNK buffer (20 mM Tris–HCl, pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20). Then 3′ends of protein-bound RNA

fragments were dephosphorylated by resuspending the beads in 20 µl PNK mix (15 µl water; 4 µl 5x PNK pH 6.5 buffer [350 mM Tris-HCl, pH 6.5; 50 mM $MgCl_2$; 25 mM dithiothreitol]; 0.5 µl PNK enzyme; 0.5 µl RNasin [Promega]) and incubating for 20 min at 37°C with shaking at 1,100 rpm. Beads were then washed once in high-salt buffer and twice with PNK buffer. Next the L3 adapter linker was ligated to RNAs 3′ends by resuspending the beads in 20 µl ligation mix (9 µl water; 4 µl 4x ligation buffer [200 mM Tris-HCl; 40 mM $MgCl_2$; 40 mM dithiothreitol]; 1 µl RNA ligase [NEB]; 0.5 µl RNasin [Promega]; 1.5 µl pre-adenylated linker L3 [20 µM]; 4 µl PEG400 [81170, Sigma]) and incubating overnight at 16°C with shaking at 1,100 rpm shaking. The next day protein-bound RNAs were radiolabelled at their 5′ends: beads were washed in high-salt buffer and PNK buffer, resupended in 8 µl of hot PNK mix (0.4 µl PNK [NEB]; 0.8 µl 32P-γ-ATP; 0.8 µl 10x PNK buffer [NEB]; 6 µl water) and incubated for 5 min at 37°C at 1,100 rpm shaking. After removal of the hot PNK mix, beads were resuspended in 20 µl 1X Nupage loading buffer (Invitrogen) and incubated at 70°C for 10 min to elute protein-RNA complexes. Supernatants were loaded on a 4-12% NuPAGE Bis-Tris gel (Invitrogen) together with a pre-stained protein size marker. The gel was run for 50 minutes at 180 V. and the protein-RNA complexes were then transferred to a nitrocellulose membrane using Novex wet transfer apparatus (Invitrogen) for 1 h at 30 V. The membranes were rinsed in PBS buffer, wrapped in saran wrap and exposed to a Fuji film at -80°C for 30 min, 1 hour or overnight.

2.13.2 iCLIP library preparation

Protein-RNA complexes were isolated from the low-RNase treated samples by cutting the membranes at the right size range (20-80 kDa above protein size) indicated by the autoradiograph. To digest the protein covalently bound to the RNAs and release peptide-RNA complexes, membranes were cut into small slices and incubated with 200 µl PK buffer (100 mM Tris-HCl pH 7.4; 50 mM NaCl; 10 mM EDTA) and 10 µl proteinase K (Roche,

03115828001) by shaking at 1,100 rpm for 20 min at 37°C. 200 µl of PK-Urea buffer (100 mM Tris-HCl pH 7.4; 50 mM NaCl; 10 mM EDTA; 7 M urea) was then added, and incubation was carried out for 20 more minutes at 37°C. The peptide-RNA complexes containing solution was added to an equal volume of RNA phenol/chloroform (acid phenol/chloroform) and added to a 2 ml Phase Lock Gel Heavy tube, incubated for 5 min at 30°C with shaking at 1,100 rpm collected by centrifugation for 5 min at 13,000 rpm at room temperature. The peptide-RNA complexes were precipitated from the aqueous solution by addition of 1:10 3M sodium acetate pH 5.5, 0.5 µl glycoblue (Ambion, 9510) and 2.5 volumes of 100% ethanol, over-night at -20°C. After washing in 80% ethanol, the primer mix (6.25 µl water; 0.5 µl RTclip primer [0.5pmol/µl]; 0.5 µl dNTP mix [10mM]) was added to the resuspended RNAs (primers listed in Table 2.10), and retro-transcription was carried out with SuperScriptIII (Invitrogen) as described (Section 2.5.2). cDNAs were precipitated in ethanol, as described earlier, resuspended in 2X TBE-urea loading buffer (Invitrogen) and then loaded on a 2X TBE-urea gel (Invitrogen) together with a low molecular weight marker. The gel was run for 40 min at 180 V and then three bands were cut at 120-200 nucleotides (high), 85-120 nucleotides (medium) and 70-85 nucleotides (low). cDNAs were recovered from the gel slices by crushing them in TE buffer with a syringe plunger and incubation with shaking at 1,100 rpm for 2 h at 37°C, then purified using Costar SpinX column (Corning Incorporated, 8161), and precipitated as described earlier.

The cDNAs were circularised by resuspending the pellets in 8 µl CircLigase mix (6.5 µl water; 0.8 µl 10x CircLigase Buffer II; 0.4 µl 50 mM MnCl2; 0.3 µl; Circligase II [Epicentre]) and incubating for 1 h at 60°C. An oligo containing the BamHI digestion site was annealed to the circularized cDNAs by adding oligo annealing mix (26 µl water; 3 µl FastDigest Buffer [Fermentas]; 1 µl cut_oligo [10 µM]) and incubating the samples for 1 min at 95°C. The temperature was decreased by 1°C/20sec until reaching 25°C, then cDNAs were digested by adding 2 µl BamHI (Fast Fermentas) and

incubating the samples for 30 min at 37°C. cDNAs were precipitated as described earlier and amplified by PCR by adding primer mix (Illumina P3/P7 primers – Table 2.10), Accuprime Supermix 1 enzyme (Invitrogen) and running the following PCR programme:

- 94°C for 2 min,
- [94°C for 15 sec, 65°C for 30 sec, 68°C for 30 sec] 28 cycles,
- 68°C for 3 min

An aliquot of the PCR product was loaded onto a precast 6% TBE gel (Invitrogen) which was then stained with Sybrgreen I DNA stain (Invitrogen) to check the extent of amplification of the three cDNA fractions of each sample before multiplexing.

Before multiplexing, excess primer was removed with AMPure XP beads (Beckam Coulter). Libraries were quantified with the Qubit® dsDNA HS Assay and High Sensitivity D1000 ScreenTape system following manufacturer's instructions.

All oligos used are listed in Table 2.10.

2.13.3 CLIP-qPCR variation

CLIP-qPCR was performed applying the following variations to the iCLIP protocol:

RNAs were not fragmented, therefore upon sonication the samples were incubated with only 2 µl TurboDNase for DNA degradation. No 3′ end adapter ligation and 5′ end radioactive labelling were perfomed.

Protein-RNA complexes were digested with proteinase K as described (Section 2.13.2) but using a final volume of 50µl. RNA was extracted from the sample as described in Section 2.5.1. cDNA synthesis and qPCR were performed as described (Section 2.5.2).

RNA enrichment was quantified as follows:

Percentage of Input (% of Input) = $100*2^{-\text{delta (Ct-IP - Ct-input)}} = 2^{-\Delta \text{ Ct}}$

Standard deviations and p-values were calculated using the Prism 7 statistical tool.

2.13.4 iCLIP bioinformatic analysis.

Reads were designed with a sample-specific adapter sequence immediately following the P7 adapter. These barcodes were 9-mers of the format NNNXXXXNN, where XXXX was a fixed 4 bp barcode for each sample (chosen to ensure adequate complexity across all samples), and the Ns were randomly incorporated bases. Samples were demultiplexed based on the 4 bp barcode allowing zero mismatches, and PCR duplicates were identified based on identical sequence of the random barcode, using an in-house perl script. Trimmed reads were mapped in single end mode to the hg38 human genome using tophat [237] with the following parameters (--no-coverage-search --max-multihits 20 --report-secondary-alignments --read-mismatches 2). Multihits (reads mapping to multi loci) were filtered, along with reads mapping with quality score less than 20.

To identify cross-linking sites, Dr Namshik Han developed a computational pipeline which identifies and analyses iCLIP cross-linking sites. The pipeline takes advantage of combining a peak-finding approach (PIPE-CLIP [PMID: 24451213]) and a differential binding detection analysis approach (dCLIP [PMID: 24398258]) using RNA-seq Input as reference and IgG peaks as background. This allows the pipeline to identify both biologically and statistically significant enriched truncation sites. The pipeline then annotates the identified cross-linking sites with a given Gencode reference annotation. The pipeline also analyses the pattern of localization and genomic features (e.g., 5'UTR, exon, intron, etc.) associated with the cross-linked sites. The genomic categories of iCLIP peak location were searched by using PAVIS [PMID: 24008416] with default settings.

To determine which regulatory motifs are over-represented in iCLIP peak regions, we identified all possible ungapped 8-mers in the regions and computed their frequency. iCLIP provides single base-pair resolution of binding position, so we needed to extend this precise peak location to allow a search of the enriched motifs. We extended +/-100 bp centred by iCLIP

peak to make 200 bp windows around the peaks. An 8-mer is considered over-represented if its frequency in the windows is significant. We found the presence of repeats that consist of a single nucleotide or dimer repeated for the entire 8-mer. This phenomenon is common in genomic sequences and generally is associated with non-functional components, and thus, these were filtered out.

To assess the statistical significance of the computed frequency for the over-represented motifs, we generated scrambled sequences according to the nucleotide composition of the original sets of sequences in the extended windows centred by iCLIP peaks. The frequencies for the random 8-mers were computed, and the distribution of the frequencies was approximated by the extreme value distribution. We used the MATLAB function *gevfit* to compute the maximum likelihood estimation of the extreme value distribution. We then overlaid a scaled version of its probability density function, computed using MATLAB function *gevpdf*, with the histogram of the frequency of the random 8-mer sequences. We repeated this process 100 times for bootstrapping and calculated the p-value. We concluded that the over-representation of the 8-mer motifs in the iCLIP peak-centered window is statistically significant if the p-value estimate is less than 1.0e-4.

To identify consensus motifs, enriched 8-mers were phylogenetically clustered into 10 groups. We used the MATLAB function *seqlinkage* to construct phylogenetic trees from pairwise distances. We checked the codon usage in the extended windows centred by iCLIP peaks. We used MATLAB *codoncount* that counts the codons in a given set of nucleotide sequences and returns the codon counts in Codons. We used the MATLAB function *seqlogo* to identify consensus motifs and their weight matrix for the clustered 8-mer(s) in each group.

The gene ontology enrichment of iCLIP bound genes was obtained using the DAVID functional annotation tool or Ingenuity pathway analysis (IPA) with default settings.

oligo	oligo sequence (5'→3')
Pre-adenylated adapter L3-App	rAppAGATCGGAAGAGCG GTTCAG/ddC/
Cut_oligo (BamHI site containing oligo)	GTTCAGGATCCACGACGCTCT TCaaaa
iCLIP RT oligo	iCLIP RT oligo sequence (5'→3')
Rt8clip (DDX3X rep1)	/5Phos/NNCATTNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rt9clip (DDX3X rep2)	/5Phos/NNGCCANNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rt11clip (DDX54 rep1)	/5Phos/NNGGTTNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rt13clip (DDX54 rep2)	/5Phos/NNTCCGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rt6clip (Rb IgG)	/5Phos/NNCCGGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rt10clip (Ms IgG)	/5Phos/NNGACCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Primer	primer sequence (5'→3')
P5 Illumina	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT
P7 Illumina	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTG CTGAACCGCTCTTCCGATCT

Table 2.9. Oligos used in iCLIP. Barcodes are highlighted.

2.14 Bioinformatic quality control analysis.

For all high throughput sequencing data (Sections 2.11.2;2.12.1;2.13.4) read quality was assessed using fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed using trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove adapter contamination.

Chapter 3 - Identification of DDX54 target RNAs through iCLIP technology

3.1 Introduction

The pathways in which DDX RNA helicases are involved and the protein complexes in which they are found largely define their overall function. However, knowing the identity of their target RNAs is essential in order to understand their specific activities and roles in cells. Moreover, mapping protein–RNA interactions is essential to dissect RNA regulatory processes. In the recent years, genome-wide tools have been developed to investigate protein–RNA interactions *in vivo*, revealing binding features of many RBPs.

The first attempts included approaches like RNA immunoprecipitation (RIP) coupled with microarrays or sequencing (RIP-Chip/RIP-Seq), which showed a series of limitations particularly regarding the high detection of non-specific interactions and the low resolution of RBP binding sites. Overcoming these limitations was the aim of developing UV cross linking-based techniques (CLIP - cross-linking immunoprecipitation). The demand of improving the resolution of this techniques is particularly important since most RBPs bind short length portions of RNAs and are therefore often attached to more than one site per RNA molecule.

UV-crosslinking a wavelength of 254 nm creates a covalent bond between an amino acid residue and an RNA base only if they are in very close proximity due to a direct interaction. Because of the UV cross linking step, CLIP-based protocols use higher stringency washes which disrupt protein–protein interactions and reduce the probability of recovery unspecific RNA. RNAs are digested to create short RBP-protected RNA fragments. This allows the identification of RBP binding sites transcriptome-wide, at high resolution. UV-crosslinking leaves an

aminoacid residue bound to the nucleotide at the protein-RNA crosslinking site. In the first version of CLIP (HITS-CLIP) the RNAs were reverse transcribed after ligation of both a 5′ adaptor a 3’ adapter and the resulting cDNAs were subjected to high-throughput sequencing. Since retro-transcription enzymes are prone to termination at the crosslinking site, it has been estimated that up to 80% of the resulting cDNA products lacked the 5′ adapter. These sequences fail to be amplified by PCR and are thus lost from the sequencing library, resulting in considerable loss of information about the binding site [245].

Individual nucleotide resolution CLIP (iCLIP) addresses the issue of low crosslinking site identification with a strategy that includes an intramolecular circularization step: reverse transcription is primed with an oligonucleotide that contains the 3′, as well as the 5′, adaptor region. The cDNA is isolated through size selection, circularized, then linearized by enzymatic digestion and amplified by PCR (Section 2.13; Figure 3.1A). Truncated cDNAs represent over 80% of the iCLIP cDNA library. Once the sequence is mapped to the genome the cross-linking site is identified by the nucleotide upstream of the truncation site. This strategy largely improved the definition of RBP binding sites [244].

iCLIP profiles can provide both biological and mechanistic insights into the function of the relevant RNA binding protein. Furthermore, when coupled with focused bioinformatic motif searches, this technique can also help to identify discrete nucleotide sequences within RNAs which are bound by the protein of interest. Since binding to a protein can also involve secondary structure of the RNAs, a bioinformatic structural prediction analysis may also reveal RNA folding patterns common to the target RNAs of a specific protein.

Although DDX helicases are related members of a broad protein family with many members thought to perform redundant functions, these proteins can show specificity with respect to the RNAs they target (Section1.3). In this chapter, and in the next, I set out to characterize the

RNA binding profiles of two members of the DDX helicase family, namely DDX54 and DDX3X. Although much is known about the cellular functions of both of these helicases, the identities of RNAs bound by these proteins were unknown at the beginning of this project.

3.2 Aims of this chapter

The aims of the work described in this chapter were:

- To determine RNA binding profiles of DDX54 in the MCF7 breast cancer cell line.
- To identify and validate individual target RNAs of the helicase.
- To identify specific RNA sequences and/or secondary structures to which the helicase binds.
- To investigate the molecular function and biological relevance of DDX54 binding to specific target RNAs.

3.3 iCLIP in-house library preparation and computational analysis of derived data

A schematic flow diagram of the iCLIP procedure is shown in Figure 3.1A. UV cross-linking and immunoprecipitation is followed by adapter ligation and radioactive labeling. When protein-RNA complexes are resolved by SDS-PAGE a radioactive smear should be visible above the molecular weight of the protein (step 8, Figure 3.1A). I performed iCLIP using highly efficient immunoprecipitation antibodies against DDX54 (Figure 3.1B) to precipitate protein-RNA complexes from UV irradiated MCF7 cells. The immunoprecipitation produced a smear within the correct size range (about 20-80 kDa above the predicted size of DDX54) (Figure 3.1C) which corresponds to the protein binding to ^{32}P-labeled RNA fragments of about 50 to 100 nucleotides, obtained by mild (+) RNase treatment. As shown in Figure 3.1C, high concentration (+++) RNase treatment leads to the

formation of protein-RNA complexes of lower molecular weights because the higher RNase activity digests protein-bound RNAs into smaller fragments. As negative controls, immunoprecipitations were performed using (i) anti IgG isotype control antibody in UV cross-linked lysates and (ii) anti-DDX54 antibodies in non-UV cross-linked lysates. For each of these conditions, no precipitation of protein-RNA complexes occurred (Figure 3.1C). This testifies to the high specificity of the protein-RNA complexes precipitated by the anti-DDX54 antibody. These complexes were used in subsequent library preparations (Section 2.13): the RNA retrieved from the DDX54 and IgG complexes was used to generate iCLIP libraries within a 150 to 300 nucleotide optimal size range (Figure 3.1D).

iCLIP detects protein-RNA binding sites at a single nucleotide resolution thanks to an intramolecular ligation step that captures cDNAs truncated at the protein binding site (steps 10 to 16 in Figure 3.1A). In order to identify these sites, we developed a computational pipeline that combines a peak-calling approach and a differential binding analysis approach (Section 2.13.4). The pipeline then annotates the identified cross-linking sites with a given Gencode reference and further analyses the pattern of localization and genomic features (e.g., 5'UTR, exon, intron, 3'UTR etc.) associated with the cross-linked sites. An enriched motif search provides an opportunity to discover specific sequential patterns of nucleotides at the cross-linked sites or in user-defined search regions. The pipeline annotates the enriched motifs with known binding motifs and generates a density plot in order to visualize the localization pattern of the motifs.
Although it would be interesting to analyse possible common structural features in all target RNAs identified by iCLIP, the comparative prediction of such a high number of targets is too demanding for the ability of available computational tools. However, it is still possible to select a subset of target RNAs, on the basis of biological common relevance, to compare discrete regions of interest surrounding the binding sites.

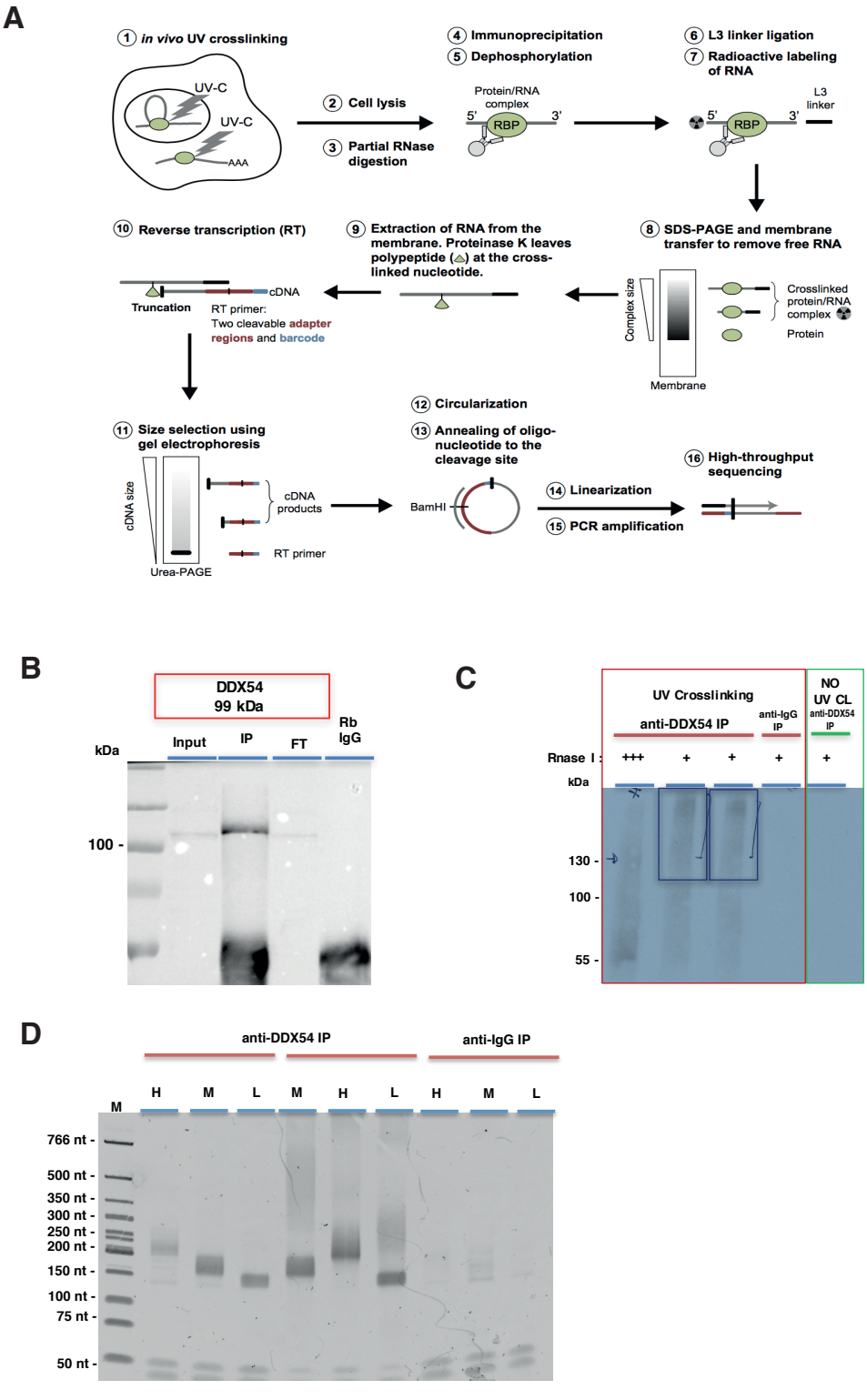


Figure 3.1. iCLIP of endogenous DDX54 in MCF7 breast cancer cells.
(A) Workflow of the iCLIP procedure from [244] (see Section 2.13 for description). (B) Western blot analysis of immunoprecipitated DDX54 in MCF7 whole cell lysate: cells were lysed in CLIP lysis buffer and DDX54 immunoprecipitation was

performed. (IP= immunoprecipitation; FT= flow through). (C) Autoradiography of DDX54 –^{32}P fragmented RNA complexes migrating 20-80 kDa above the size of the protein (indicated by the blue box in the second and third lanes). Whole lysates from MCF7 cells exposed to 254 nm UV crosslinking at 150mJ/cm^2 were subjected to immunoprecipitation using an anti-DDX54 antibody in the presence of high RNase (+++) treatment (single sample), or low (+) RNase treatment (in duplicate) or anti-IgG antibody immunoprecipitation in the presence of mild RNase treatment (+) (single sample). Whole lysate from MCF7 cells not exposed to UV crosslinking was used for anti-DDX54 immunoprecipitation in the presence of low (+) RNase treatment (single sample). (D) DDX54 iCLIP libraries: RNA was recovered from immunoprecipitated material and size-selected DDX54:RNA complexes shown in (C). cDNA was divided into three size ranges and libraries were produced as described (Section 2.13.2). Shown are the PCR amplified libraries obtained for each size range for both biological replicates of DDX54 immunoprecipitation and the single IgG control immunoprecipitation (H=high; M=medium; L=low nucleotide size range).

Using our pipeline, we identified numerous DDX54 target RNAs and some potential protein binding sites within their sequences (Appendix Figure 1). These results are presented and discussed in the following sections.

3.4 DDX54 binds to introns of nuclear protein mRNAs and to non-coding RNAs

As shown in Figure 3.2A-C, bioinformatic analysis of the DDX54 iCLIP showed localization of truncation sites within introns towards the 3′ end of coding transcripts as well as in non-coding RNAs, including long non-coding RNAs. To validate these DDX54 binding data, I performed CLIP-qPCR on four putative targets (*NEAT1, PVT1, TET2 and NCOA1:* Figure 3.3) and one negative control unbound mRNA (*PUS1*). CLIP-qPCR was performed with an anti-DDX54 antibody, as well as with antibodies against others DDX helicases and an IgG control. Target RNAs were specifically and significantly enriched by DDX54 antibody and the control non-binding RNA, *PUS1,* was unspecific in DDX54 CLIP-qPCR (Figure 3.3).

A DAVID gene ontology analysis of DDX54 target RNAs showed enrichment for nucleoplasmic components with transcriptional or chromatin related functions (Figure 3.4A-B). Cell fractionation analysis and immunofluorescence assay showed DDX54 to be localized in the nucleolus and in the nucleoplasm of MCF7 cells (Figure 3.5A-B). Results of pathways analysis highlighted that many DDX54 target RNAs are part of “upstream regulators” enriched groups, such as Estrogen Receptor alpha (ESR1) and p53 (TP53) (Figure 3.4C).

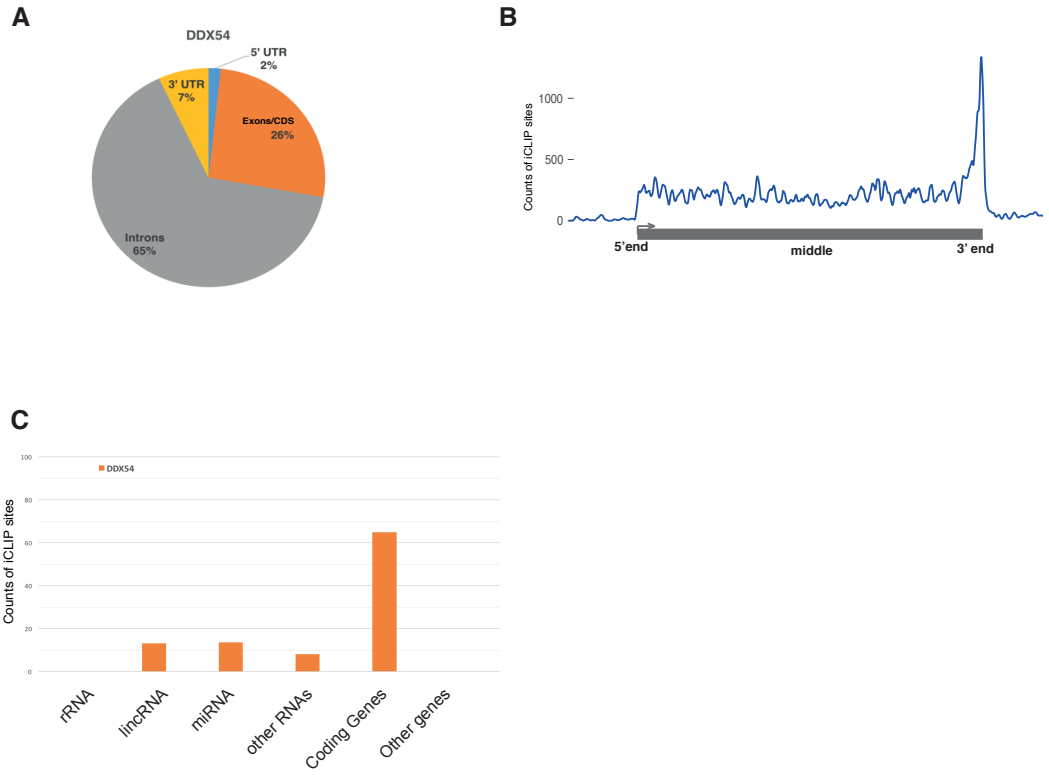


Figure 3.2. DDX54 iCLIP truncation sites localize preferentially to intronic regions and at the 3'end of transcripts.
(A) Pie chart indicating location of iCLIP peaks in transcripts. Shown are the percentages of DDX54 iCLIP truncation sites found in the indicated category. The genomic categories of iCLIP peak location were searched for using PAVIS [PMID: 24008416] with default settings.
(B) Density plot of iCLIP localisation on a scaled gene body (metaplot). The transcripts that have iCLIP peaks of DDX54 were scaled to the same length and also extend by 20% up- and down-stream. Then, iCLIP peaks for each gene were accumulated on the scaled gene to visualise the distribution of iCLIP truncation sites localisation. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis = counts of iCLIP sites; X axis = scaled gene body localization. (C) Bar graph of gene types containing iCLIP peaks. Shown are the number of DDX54 iCLIP truncation sites in each gene type category. Y axis = counts of iCLIP sites; X axis = gene type category.

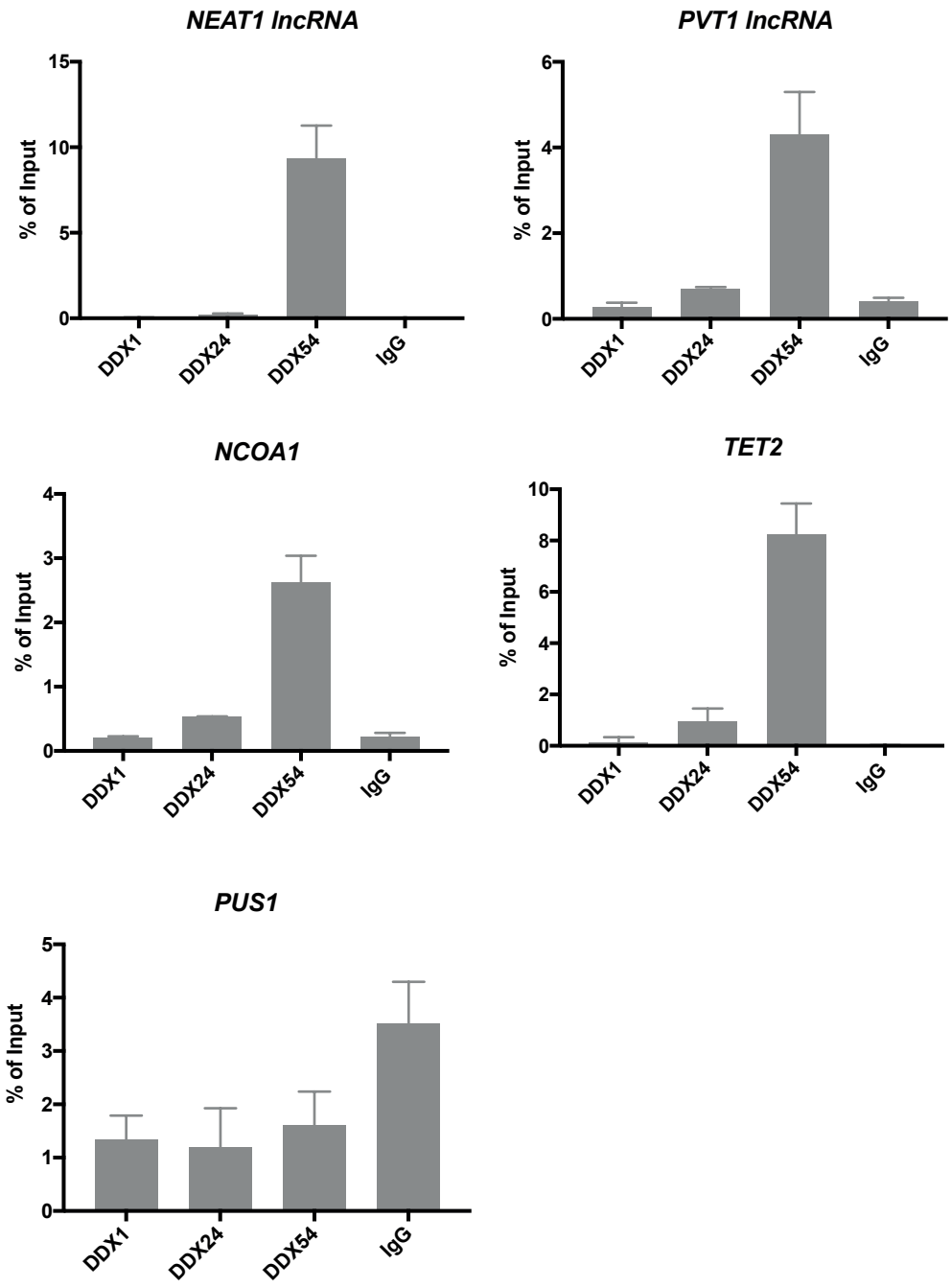
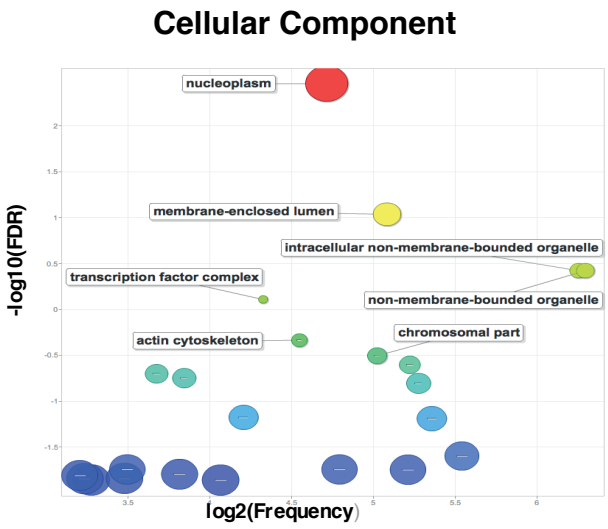


Figure 3.3. CLIP-qPCR validation of DDX54 iCLIP targets.
CLIP-qPCR was performed as described (Section 2.13.3) to validate the binding of *NEAT1 lncRNA, PVT1, NCOA1*, *TET2* and *PUS1* mRNAs to DDX54. Antibodies against DDX1 and DDX24, and against IgG, were controls, as was analysis of *PUS1,* an RNA not identified as binding DDX54. cDNA obtained from each immunoprecipitation was analysed by qPCR to test for enrichment of the indicated targets. Results represent the average of three biological replicates. Error bars indicate standard deviation.

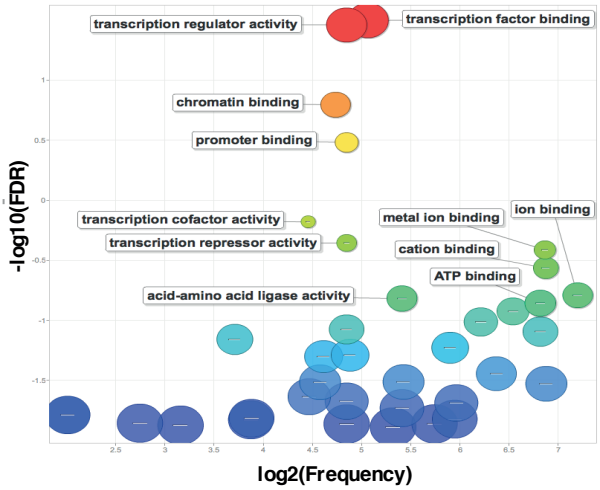
I next analyzed the sequence features of DDX54 binding sites and observed enrichment of A- and T- (i.e. U- in RNA) rich triplets in both 200 base-pair and 30 base-pair windows centred around truncation sites (Figure 3.6). To investigate whether these triplets could be part of a sequence binding motif, I performed a motif search analysis and identified ten subgroups of related motifs to which most DDX54 targets could be assigned (Figure 3.7A). The motif search surrounding DDX54 binding sites showed a good extent of overlap between the motifs and the localization of the truncation sites (Figure 3.7B), which suggests that DDX54 most likely binds to target RNAs via specific motifs.

Examples in Figure 3.8A-C show the enrichment of poly A and poly T (i.e. poly U in RNA) sequences at the RNA binding sites of DDX54, which was predicted already by the triplet usage analysis (Figure 3.6). Among the motifs found, one T-rich (i.e. U-rich in RNA) motif showed homology with the splicing factors U2AF2 and USAF50 RNA recognition sequence (Figure 3.8C). This is consistent with a role for DDX54 in splicing, recently shown by others [183].

A


Cellular Component
nucleoplasm
membrane-enclosed lumen
intracellular non-membrane-bounded organelle
transcription factor complex
non-membrane-bounded organelle
actin cytoskeleton
chromosomal part
-log10(FDR)
log2(Frequency)

B


Molecular Function
transcription regulator activity
transcription factor binding
chromatin binding
promoter binding
transcription cofactor activity
transcription repressor activity
metal ion binding
ion binding
cation binding
ATP binding
acid-amino acid ligase activity
-log10(FDR)
log2(Frequency)

C

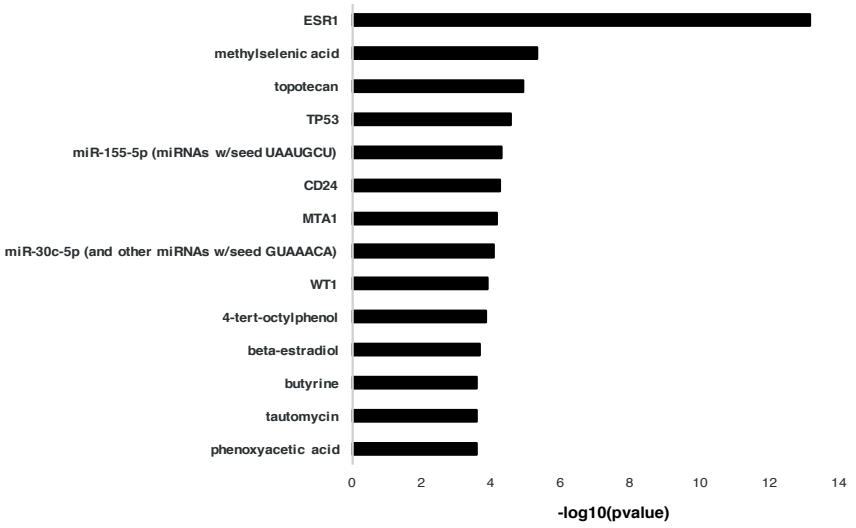

ESR1
methylselenic acid
topotecan
TP53
miR-155-5p (miRNAs w/seed UAAUGCU)
CD24
MTA1
miR-30c-5p (and other miRNAs w/seed GUAAACA)
WT1
4-tert-octylphenol
beta-estradiol
butyrine
tautomycin
phenoxyacetic acid
0
2
4
6
8
10
12
14
-log10(pvalue)

Figure 3.4. Gene ontology analysis of DDX54 iCLIP target RNAs.
DAVID gene ontology (GO) search of cellular components (A) and molecular function (B) enriched terms. The GO enrichment of DDX54 iCLIP bound transcripts was obtained using the DAVID functional annotation tool with default settings [PMID: 19131956]. P-values were calculated using the "default" method of DAVID.
(A-B) The scatter plot was generated with REVIGO [PMID: 21789182] to visualise the GO term analysis result for cellular components (A) and molecular function (B). The scatter plot shows GO term cluster representatives in a two-dimensional space. X-axis indicates the frequency of the GO term in the underlying GO analysis database (i.e. bubbles that are closer to the end of x-axis (lager values) are more general terms). Y-axis and bubble colour/size indicates the log transformed p-value of GO terms.
(C) Bar graph shows Upstream regulators term enrichment generated with IPA tool (IPA (Ingenuity® Systems) GO search of DDX54 iCLIP targets. P-values were calculated using default settings of Ingenuity. The Y-axis shows the GO terms related to upstream regulators. X-axis shows –log10 p-values of each GO term. The group with highest significance was ESR1 (Estrogen Receptor α).

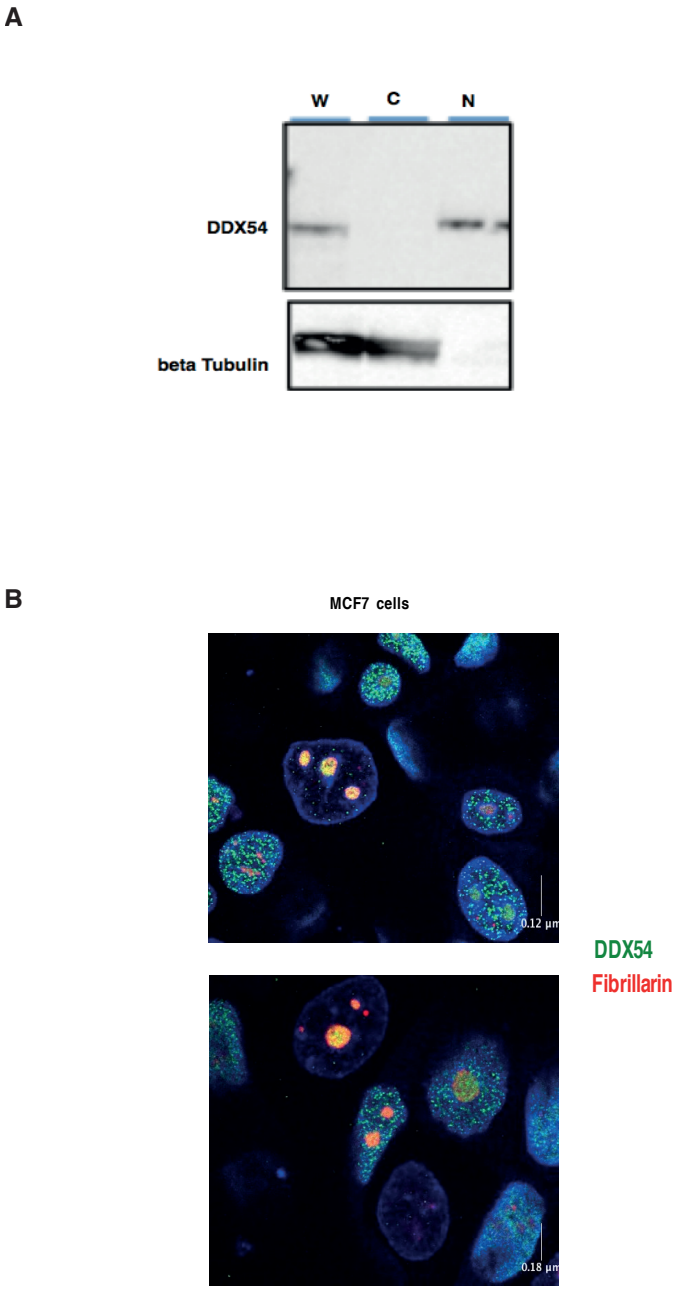


Figure 3.5. DDX54 is localized in the nuclei of MCF7 cells.
(A) Western blot detection of DDX54 and β-tubulin in whole cell (W), cytoplasmic (C) and nuclear (N) fractions of MCF7 cells.
(B) DDX54 immunofluorescence in MCF7 cells. Cells were fixed and probed with anti-DDX54 mouse antibody (ab57711) and anti-Fibrillarin rabbit antibody. Nuclei were stained with DAPI; DDX54 was detected with Alexa488-conjugated goat anti-mouse IgG2A; Fibrillarin was detected with Alexa594-conjugated goat anti-rabbit IgG. Scale bar length: 50 pixel; Scale bar width: 0.12 µm or 0.18 µm (as indicated).

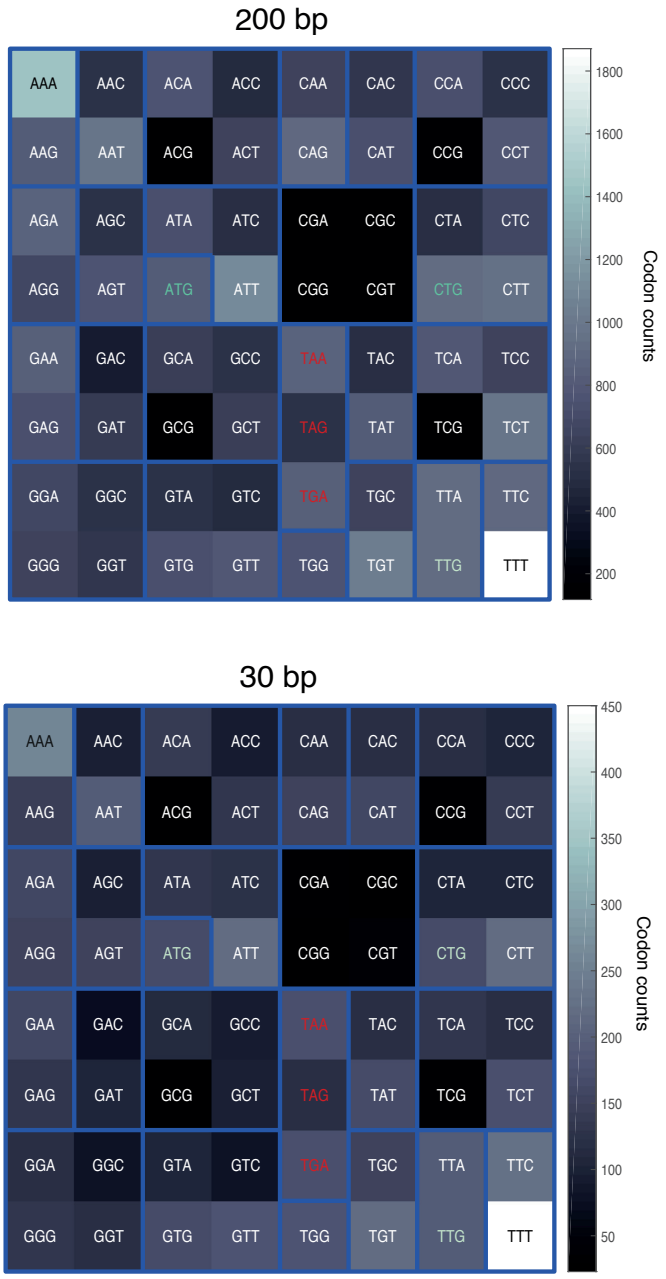


Figure 3.6. Motif search on DDX54 iCLIP binding sites: a poly A binder.
Heatmap of triplet usage in 200 base-pair and 30 bp windows centered by DDX54 iCLIP peak. Shown is the enrichment of triplets found in a 200 bp (upper panel) and 30 bp (lower panel) extended window centred on iCLIP peaks. MATLAB *codoncount* was used to generate the heatmap (Section 2.13.5). Scale indicates counts of triplets found in the considered window and associated hub.

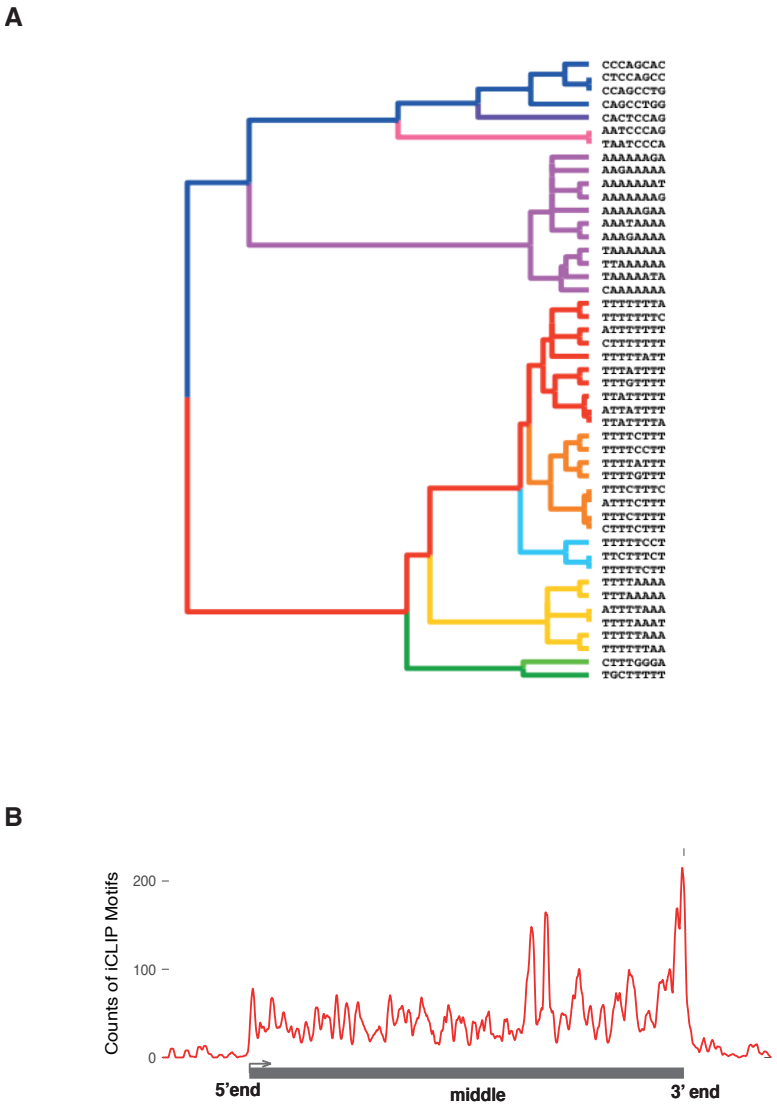
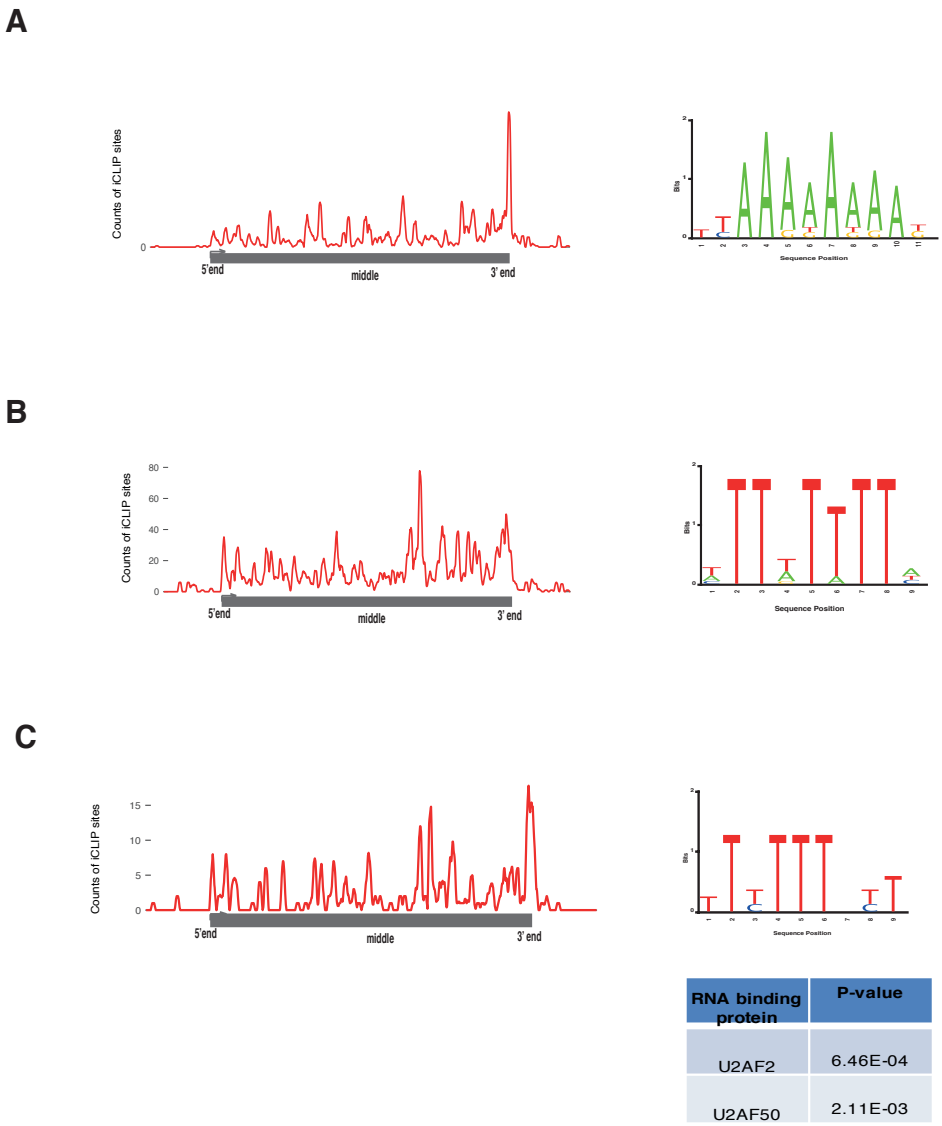


Figure 3.7. Motif search on DDX54 iCLIP binding sites.
(A) Phylogenetic tree of highly enriched 8-mer motifs in 200 base-pair windows centered on DDX54 iCLIP peak. 8-mers enriched in an extended +/-100bp centred by iCLIP were found as (Section 2.13.5). To identify consensus motifs, enriched 8-mers were phylogenetically clustered into 10 groups. MATLAB function *seqlinkage* was used to construct the phylogenetic tree shown (Section 2.13.5). (B) Density plot of highly enriched motifs localised within a scaled gene body. The distribution of the motifs within a scaled gene are shown. The transcripts that have the enriched motifs in the +/-100bp extended windows centred by DDX54 iCLIP peaks were scaled to the same length and also extended by 20% up- and down-stream. Then, the enriched motifs for each gene were accumulated on the scaled gene to visualise the distribution of DDX54 binding motif localisation. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis = counts of iCLIP motifs; X axis = scaled gene body localization.



RNA binding protein	P-value
U2AF2	6.46E-04
U2AF50	2.11E-03

Figure 3.8. Examples of enriched motifs at DDX54 iCLIP binding sites.
Enriched motifs at DDX54 binding sites. Density plot of a motif's location within a scaled gene body. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis indicates the frequency (actual count) of the motif; X axis = scaled gene body localization. Consensus motifs for the clustered 8-mer(s) in each phylogenetic group were found with MATLAB function *seqlogo* (Section 2.13.5). The sequence logos of two different phylogenetic groups (A and B) are shown (Y axis = Bits; X axis =sequence position). (C) The motif is matched to known RNA binding motif. The table shows the known RNA binding proteins that recognize a similar sequence motif and relative p-values indicating the significance of homology.

3.5 Investigations into DDX54 binding NEAT1 lncRNA

NEAT1 lncRNA was one of the strongest DDX54 validated targets (Figure 3.3). I further confirmed the specificity of this reaction by performing CLIP-qPCR analysis in MCF7 cell extracts depleted of DDX54 using shRNA (Figure 3.9A-B).

NEAT1 lncRNA is localized within discrete subnuclear structures called paraspeckles [246]. I therefore investigated whether binding of DDX54 affects formation of *NEAT1* lncRNA-dependent paraspeckles. For this purpose, I performed RNA FISH to localize *NEAT1* RNA in MCF7 cells transfected with either scrambled siRNA or two different DDX54 specific siRNAs. DDX54 protein level was efficiently reduced by both siRNAs (#9 and #12) (Figure 3.10A), but no significant depletion of paraspeckles was observed in MCF7 cells transfected with these siRNAs, compared to MCF7 cells transfected with scrambled siRNA (Figure 3.10C).

Importantly, by combining RNA FISH for *NEAT1* lncRNA detection and immunofluorescence for DDX54 detection, I did not observe co-localization of DDX54 and *NEAT1* lncRNA in wild type MCF7 cells (Figure 3.10D). Also, the expression levels of *NEAT1* lncRNA was not consistently affected by DDX54 knock down using the two different siRNAs (Figure 3.10B). This discrepancy could be partially explained by the different efficacy of the two siRNAs. In fact, the higher depletion of DDX54 when siRNA #12 was used led to a significant change in *NEAT1* lncRNA levels, not observed with siRNA #9-mediated DDX54 knock down (Figure 3.10A-B). Taken together, these data do not support my hypothesis that DDX54 is involved in formation of NEAT1 paraspeckles.

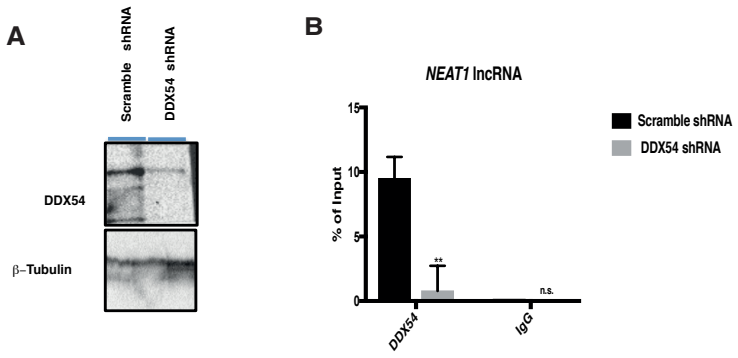


Figure 3.9. NEAT1 lncRNA binding validation in DDX54 knock down MCF7 cells.

(A) Western blot shows the levels of DDX54 and ß-tubulin in MCF7 cells transduced with scrambled or DDX54 targeting shRNA, 72 h after transfection. (B) DDX54 and IgG CLIP-qPCR were performed as described (Section 2.13.3) in MCF7 cells transduced with scrambled or DDX54 targeting shRNAs. Results represent the average of three biological replicates. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble shRNA (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

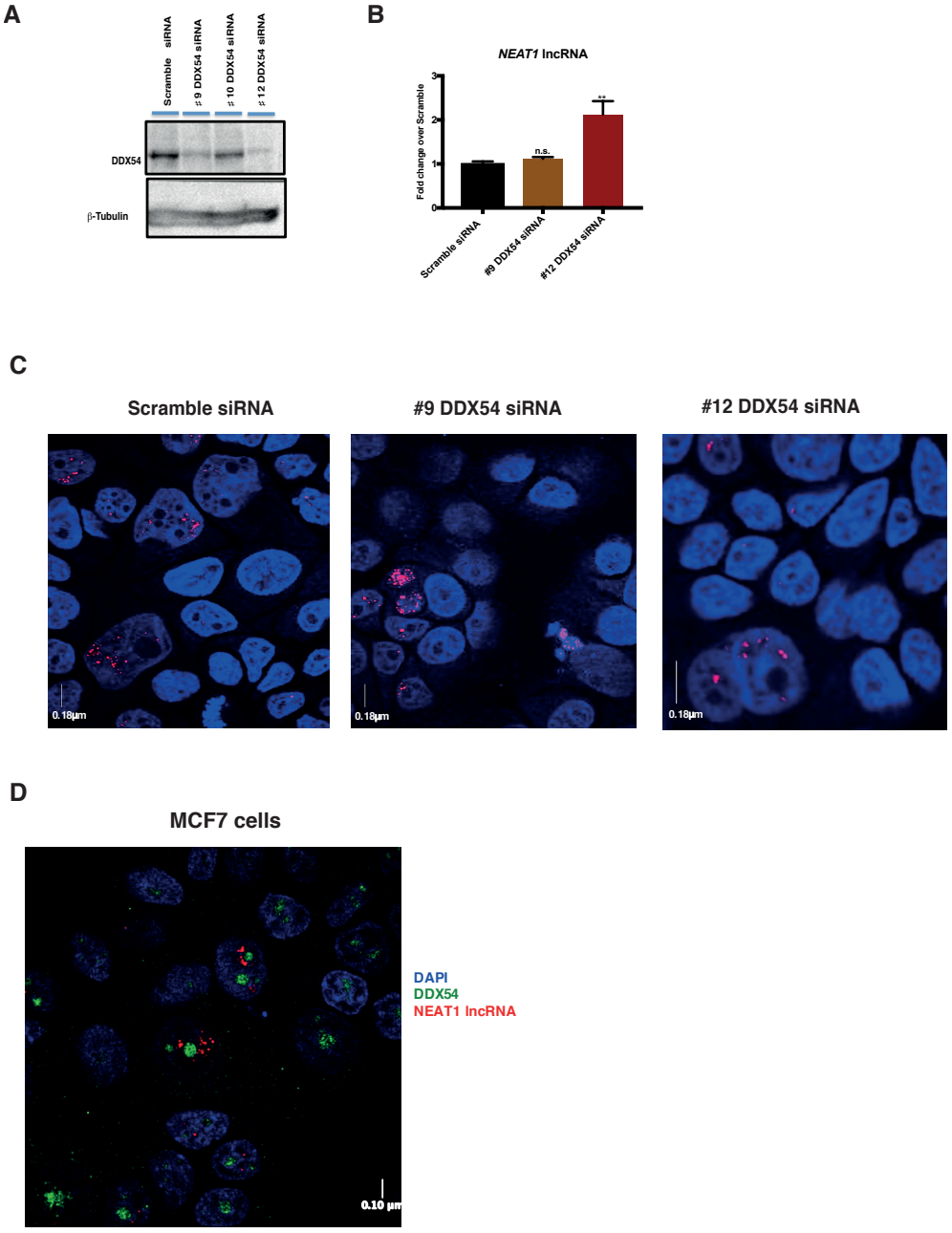


Figure 3.10. Knock down of DDX54 does not affect *NEAT1* lncRNA.
(A) A western blot shows the level of DDX54 and ß-tubulin in MCF7 cells transfected with either scrambled siRNA or three different DDX54 targeting siRNAs (#9, #10 or #12), alternatively.
For all conditions, cells were harvested 72h after transduction/transfection. (B)*NEAT1* lncRNA levels in MCF7 cells transfected with either scramble siRNA or two different DDX54 targeting siRNAs (#9 or #12), measured by qPCR. Results represent the average of three biological replicates (Section 2.5.3). Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001). (C) RNA FISH was performed as described (Section 2.9.2) in MCF7 cells transfected with either scramble siRNA or two different DDX54 targeting siRNAs (#9 or #12). Images show cell nuclei

stained with DAPI. Human *NEAT1* lncRNA stellaris probes were conjugated with Quasar 570 fluorescent dye. (D) DDX54 immunofluorescence and *NEAT1* lncRNA FISH in wild type MCF7 cells. Cells were fixed as described (Section 2.9.3) and probed with anti-DDX54 mouse antibody (ab57711) and anti-Fibrillarin rabbit antibody. Nuclei were stained with DAPI; DDX54 was detected with Alexa488-conjugated goat anti-mouse IgG2A; Fibrillarin was detected with Alexa594-conjugated goat anti-rabbit IgG.
Scale bar length: 50 pixels; Scale bar width: 0.18 μm or 0.10 μm (as indicated). All images were taken with an Olympus FV1000 upright microscope.

3.6 Contradictory evidence of DDX54's Oestrogen Receptor α co-transcriptional regulation activity

As previously described (Section 1.4), DDX54 is a co-transcriptional regulator of several nuclear receptors. In particular, it co-represses estrogen receptor α (ERα)-dependent transcription [180]. My gene ontology analysis of DDX54 iCLIP target RNAs showed significant enrichment of genes belonging to the ERα pathway (ESR1 upstream regulator - Figure 3.4C). I therefore hypothesized that DDX54 exerts its regulatory activity by co-transcriptionally binding to RNAs transcribed from ERα target genes. To test this hypothesis, I performed ChIP-seq experiments to determine the genomic locations of ERα and DDX54 in MCF7 cells (Appendix Figure 2). ERα ChIP-seq analysis revealed regions bound by ERα, which were predominantly promoters, whilst DDX54 ChIP-seq analysis failed to identify any significant binding events. This is exemplified in Figure 3.11A that shows ChIP-seq genome tracks surrounding two ERα target gene promoters (*TFF1* and *GREB1*). It highlights the absence of peaks in the DDX54 ChIP-seq track. I also performed co-immunoprecipitation experiments to investigate whether DDX54 and ERα interact in MCF7 cells. Immunoprecipitation of ERα resulted in DDX54 pull down but I could not observe the reciprocal pull down of ERα in DDX54 immunoprecipitation, even with a longer exposure (Figure 3.11B).

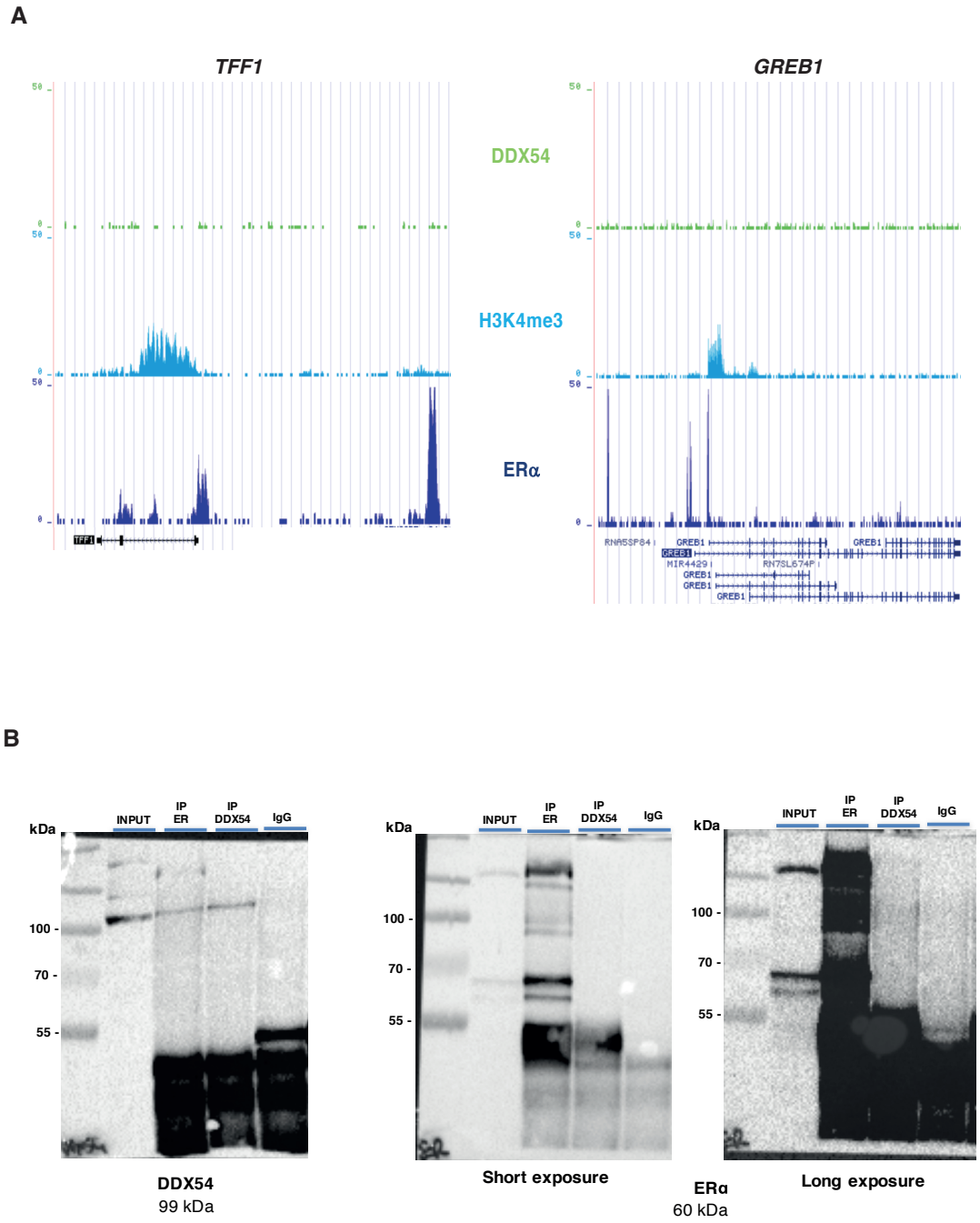
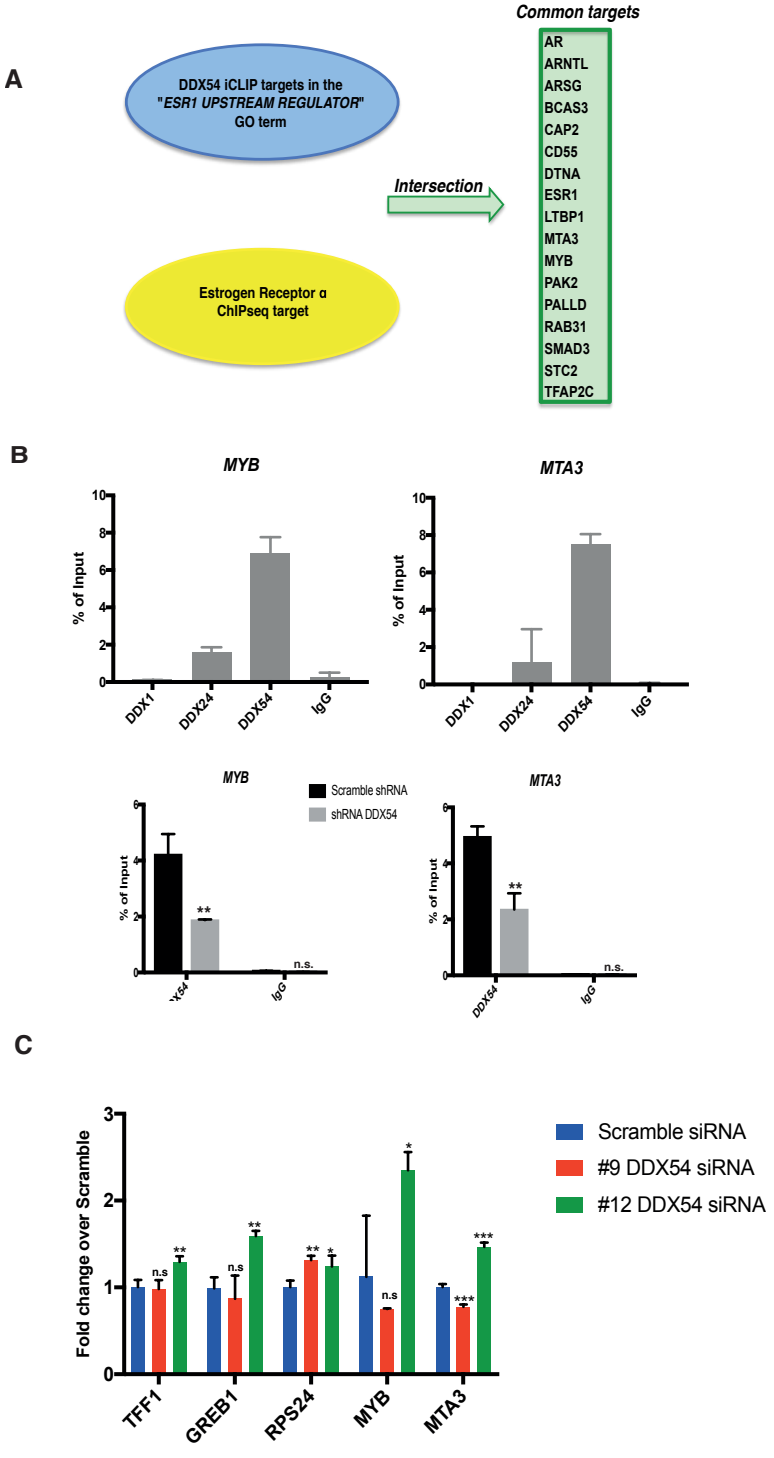


Figure 3.11. Lack of evidence for DDX54 and ERα co-localization at promoters. (A) USCS custom track for *TFF1* and *GREB1* genes showing ChIP-seq binding data for ERα, DDX54 and H3K4me3. H3K4me3 is seen interspersed over the 5' side of the gene body (*TFF1*) and towards the promoter (*GREB1*). ERα binding is enriched at promoter regions of both genes. DDX54 track shows no peaks. The gene is represented below the X-axis and the height of the ChIP-seq read density is annotated on the Y-axis and is manually scaled at the same height for all tracks. (B) Co-immunoprecipitation of oestrogen receptor α and DDX54 performed as described (Section 2.6.5).


A
DDX54 iCLIP targets in the "ESR1 UPSTREAM REGULATOR" GO term
Estrogen Receptor α ChIPseq target
Intersection
Common targets
AR
ARNTL
ARSG
BCAS3
CAP2
CD55
DTNA
ESR1
LTBP1
MTA3
MYB
PAK2
PALLD
RAB31
SMAD3
STC2
TFAP2C
B
MYB
MTA3
% of Input
DDX1
DDX24
DDX54
IgG
Scramble shRNA
shRNA DDX54
n.s.
C
Fold change over Scramble
Scramble siRNA
#9 DDX54 siRNA
#12 DDX54 siRNA
TFF1
GREB1
RPS24
MYB
MTA3

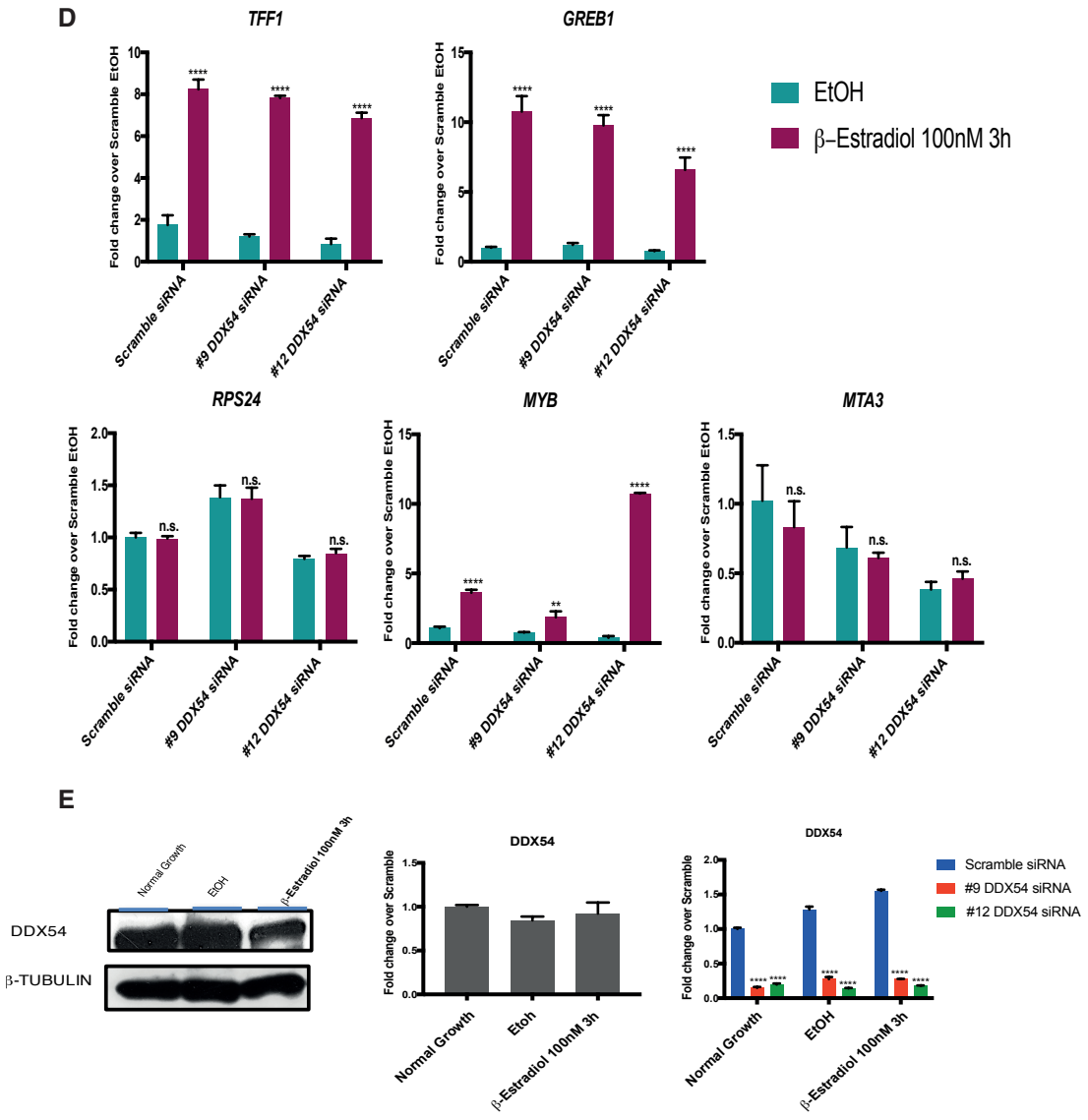


Figure 3.12. DDX54 does not regulate ERα dependent transcription.
(A) ERα ChIP-seq and DDX54 CLIP targets belonging to the "ESR1 upstream regulator" GO term were intersected. Shown is the list of common targets between the two data sets. (B) CLIP-qPCR was performed as described (Section 2.13.3) to validate binding of *MTA3* and *MYB* mRNAs to DDX54, and to DDX1, DDX24 and IgG as controls (top panels). DDX54 and IgG CLIP-qPCR were performed in MCF7 cells transduced with scrambled or DDX54 targeting shRNA (bottom panels). (C) mRNA levels of *TFF1, GREB1, RPS24, MYB and MTA3* in MCF7 cells transfected with either scramble siRNA or two different DDX54 targeting siRNAs (#9 or #12), under normal growth conditions. (D) mRNA levels of *TFF1, GREB1, RPS24, MYB* and *MTA3* in MCF7 transfected either with scrambled siRNA or two different DDX54 targeting siRNAs (#9 or #12). Cells were grown in estrogen deprived medium for 72h and treated for 3 hours with either 100% ethanol or 100nM β-estradiol. Results represent the average of three biological replicates (Section 2.5.3). (E) Western blot (left panel) shows DDX54 and β-tubulin proteins levels in MCF7 cells grown in normal condition or in

estrogen deprived medium for 72h and treated for 3 hours with either 100% ethanol or 100nM β–estradiol; DDX54 mRNA levels in the same conditions in cells (middle panel). The right panel shows DDX54 mRNA levels measured in qPCR: cells were treated as in D. Results represent the average of three biological replicates (Section 2.5.3). Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble siRNA (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

These latest results suggested that DDX54 regulation of ERα activity does not occur at the chromatin level. Nevertheless, DDX54 binding to ERα target genes' RNA may be involved in the stability or processing of such transcripts. Indeed, this form of regulation may be specifically relevant to the biology of MCF7 breast cancer cells. To test whether DDX54 is required for an efficient and effective response to activation of ERα dependent transcription, MCF7 cells were grown in oestrogen-deprived media for 72h and then oestrogen receptor signaling was reactivated by addition of β-estradiol to the culturing media. This system allows analysis of early oestrogen responsive genes thereby distinguishing real targets of ERα transcriptional regulation from secondary events.

Based on published observations that DDX54 co-represses *TFF1* (one of the main ERα early responsive genes), in oestrogen-stimulated transcription [180], I investigated whether DDX54 has a general role in regulating ERα target genes. In order to expand the scope of my analysis, I included other putative ERα target genes. I intersected the results from the ERα ChIP-seq with those from the DDX54 iCLIP to identify DDX54 target RNAs that are also transcriptional ERα targets (Figure 3.12A). Among the genes highlighted by the CLIP/ChIP intersection, I selected two candidates to test this hypothesis: *MYB* and *MTA3*. Figure 3.12B shows validation by CLIP-qPCR of DDX54 binding to *MYB* and *MTA3* RNAs, performed as previously described in wild type MCF7 cells and in MCF7 cells transduced with either scrambled shRNA or DDX54 targeting shRNA. I then measured expression of two known ERα transcriptional targets (*TFF1* and *GREB1)*, *MTA3* and *MYB* and *RPS24* (which is not an ERα ChIP target and was therefore used as a negative control). In MCF7 cells under normal growth conditions, *TFF1, GREB1, MYB* and *MTA3* expression increased to different extents in DDX54 knock down only when one of two DDX54 siRNAs was used. *RPS24* expression was slightly increased in DDX54 knock down using both siRNAs (Figure 3.12C).

When I tested the same genes in conditions of ERα signalling early activation, I found the two main ERα target genes *TFF1* and *GREB1* were

consistently up-regulated upon β-estradiol treatment and their level of β-estradiol dependent activation showed no change, or only a slight reduction, when DDX54 was knocked down with either of the two siRNAs (Figure 3.12D). As expected, *RPS24* expression did not change irrespective of the presence of β-estradiol in the growth medium, whereas *MYB* and *MTA3* did not show consistency in their response to neither β-estradiol dependent ERα activation nor when comparing the response obtained with either of the two siRNAs used to knock down DDX54 in these conditions (Figure 3.12D). Figure 3.12E shows the level of *DDX54* mRNA and protein did not change upon growth in oestrogen-deprived medium nor during the β-estradiol treatment.

Overall, these findings suggest DDX54 does not co-regulate transcriptional activity of ERα in MCF7 cells.

3.7 Discussion

DDX54 iCLIP in MCF7 cells was performed following successful quality controls steps, as determined by published protocols (Figure 3.1B-D). The data generated were used in the development of our iCLIP data analysis pipeline. This analysis uncovered some relevant information about the RNA binding preferences of DDX54 in MCF7 cells. Nevertheless, downstream investigations into the biological role of this helicase were inconclusive, probably also due to the low correlation of the iCLIP replicates generated (Appendix Figure 1B). The results of this chapter also highlight the need to couple iCLIP data to other data sets such as those derived from RNA-seq, ribosome profiling or mutant studies, to lead to a more comprehensive overview of potential functions of a protein of interest.

3.7.1 DDX54 shows binding preferences toward A-rich introns and 3' ends

One important finding from bioinformatic analysis of DDX54 iCLIP was that the DDX54 binding sites preferentially localize toward the 3′ end of target transcripts and correspond to A- and T- (i.e. U- in RNA) rich triplets. These observations would suggest preferential binding to mRNA tails. Given the fact that DDX54 is a nuclear protein, it is unlikely that poly A binding would be related to translational processes. One speculation would be that DDX54 could be involved in polyadenylation events during or just after transcription. DDX54 was also found to bind preferentially to introns. Additionally, one T- (i.e. U- in RNA) rich sequence among the enriched motifs found in DDX54 binding sites, suggested a probable role in splicing. In fact, this sequence shows homology to the RNA sequence recognized by splicing factors (U2AF2, U2AF50).

The DAVID Gene Ontology analysis of DDX54 target RNAs showed enrichment for nucleoplasmic components with transcriptional or chromatin related functions. This is an interesting observation if linked to the subcellular localization of DDX54. I indeed observed that DDX54, which is uniquely localized in the nucleus, showed preferential binding to nuclear related proteins RNAs. These observations would support the idea that proteins regulating RNAs of factors functional in a particular subcellular compartment are localized in the same relevant region.

DDX54 is one of the least characterized DDX helicases, however a few reports in the literature indicate a role in regulating transcriptional activity of nuclear receptors (Section 1.4). Taking advantage of these published observations, I tried to expand the supposed ERα co-regulatory role of DDX54 in MCF7 breast cancer cells. However, my results did not show consistent evidence for a regulatory role of DDX54 on ERα-dependent transcription. *TFF1* and *GREB1* are ERα target genes that respond early to oestrogen mediated activation of the receptor. These genes were not included in the overlap of ERα ChIP-seq targets and

DDX54 iCLIP targets. Moreover, I observed inconsistent effects of both DDX54 knock down and β-estradiol dependent ERα activation on the expression of genes that resulted from the overlap of the two data sets (*MYB* and *MTA3*). Altogether, these observations suggest DDX54 does not affect ERα-dependent transcription and that ERα transcriptional target genes are enriched as DDX54 iCLIP targets as a consequence of their expression in an ERα-dependent breast cancer cells line (MCF7) but with no apparent biological relevance for DDX54 binding their mRNAs.

Chapter 4 - Identification of DDX3X target RNAs through iCLIP technology

4.1 Introduction

The aim of Chapter 3 was to identify DDX54 RNA binding targets in order to gain novel insights into the biological function of a so-far poorly characterised protein. Unfortunately, the outcome of the DDX54 iCLIP analysis and the resulting follow-up investigations did not provide particularly useful information concerning the function of DDX54 on its target RNAs. As this outcome was considered a possibility, and in order to widen the scope of my investigation, I also performed in parallel an iCLIP experiment in MCF7 cells on a second candidate DEAD-box RNA helicase: DDX3X.

As discussed in Section 1.5, DDX3X is a more well characterized protein than DDX54, and it has been shown to perform several molecular functions in post-transcriptional events. Although the role of DDX3X in breast cancer *in vitro* models has been partially investigated (Section 1.6.1), the RNA target-specific activity of this protein in breast cancer cells was unexplored.

4.2 Aims of this chapter

Similar to the previous chapter, the aims of the work described in this chapter were:

- To determine the RNA binding profiles of DDX3X in the MCF7 breast cancer cell line.
- To identify and validate individual target RNAs.
- To identify specific RNA sequences and/or secondary structures to which the helicase binds.

- To investigate molecular function and biological relevance of DDX3X binding to specific target RNAs.

4.3 DDX3X iCLIP in-house library preparation

Similarly to that shown for DDX54 in the previous chapter, I performed iCLIP using highly efficient immunoprecipitation antibodies against DDX3X (Figure 4.1A) to precipitate protein-RNA complexes from UV irradiated MCF7 cells. Figure 4.1B show the smear produced by DDX3X binding to ^{32}P-labeled RNA fragments bound to DDX3X. These fragments of about 50 to 300 nucleotides, were obtained upon mild (+) RNase treatment of the RNA cross-linked to DDX3X and following ^{32}P-labelling. As with DDX54 immunoprecipitation, high concentration (+++) RNase treatment leads to the formation of protein-RNA complexes of lower molecular weight. These give a signal just above the size of DDX3X due to the protein binding short RNA fragments. Negative control immunoprecipitations with (i) anti IgG isotype control antibody in UV cross linked lysates or (ii) anti-DDX3X antibodies in non-UV cross linked lysates did not precipitate any protein-RNA complexes (B). Protein-RNA complexes precipitated by the anti-DDX3X and anti-IgG antibodies were then used for subsequent library preparation (Section 2.13). RNA was retrieved from the complexes and used to generate iCLIP libraries within a 150 to 300 nucleotide optimal size range (Figure 4.1C)

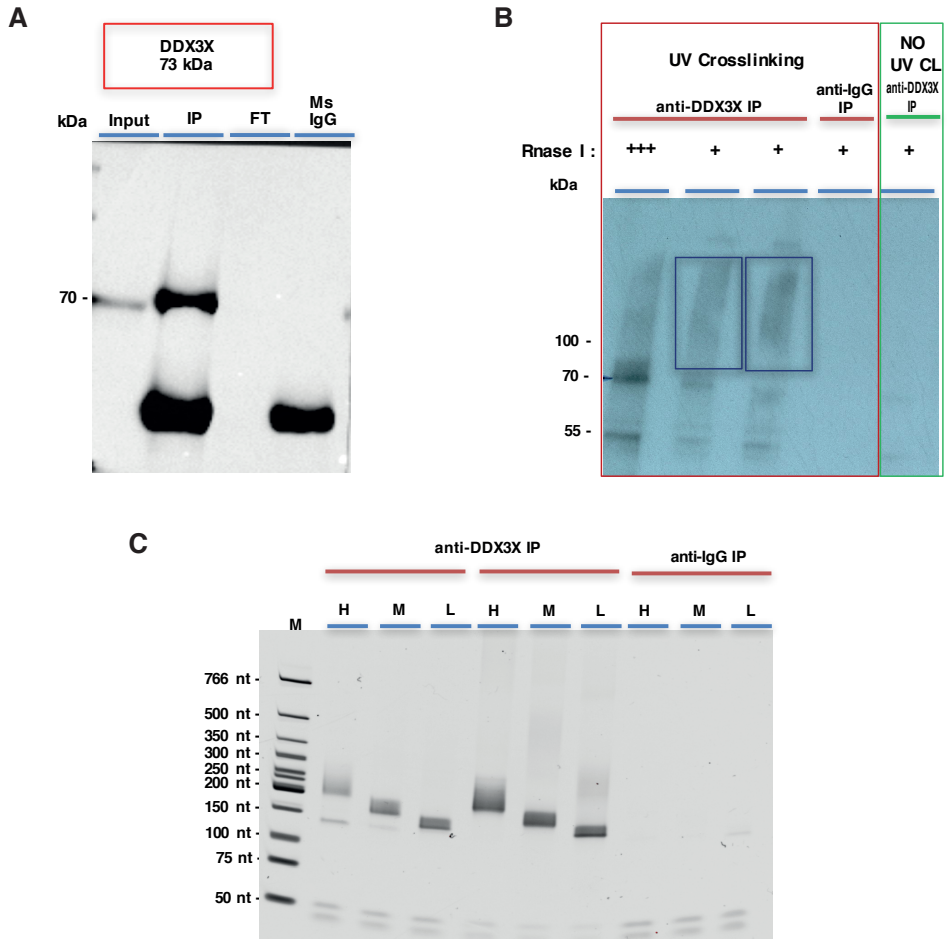


Figure 4.1. iCLIP of endogenous DDX3X MCF7 breast cancer cells.
(A) Western blot analysis of immunoprecipitated DDX3X in MCF7 whole cell lysate: cells were lysed in CLIP lysis buffer and DDX3X immunoprecipitation was performed. (IP= immunoprecipitation; FT= flow through). (B) Autoradiography of DDX3X - ^{32}P fragmented RNA complexes migrating 20-80 kDa above the size of the protein (indicated by the blue boxes in the second and third lanes). Whole cell lysates from MCF7 cells exposed to 254 nm UV crosslinking at 150mJ/cm^2 were used to immunoprecipitate DDX3X in the presence of high (+++) RNase treatment in a single sample and low (+) RNase treatment in duplicate and anti-IgG antibody immunoprecipitation in the presence of mild RNase treatment in a single sample. Whole lysate from MCF7 cells not exposed to UV was used for anti-DDX3X immunoprecipitation in the presence of low (+) RNase treatment in a single sample. (C) DDX3X iCLIP libraries: RNA was retrieved from immunoprecipitated complexes and size-selected DDX3X-RNA complexes shown in (B). cDNA was divided into three size ranges and libraries were produced as described (Section 2.13.2). Shown are the PCR amplified libraries obtained for each size range for both biological replicates of DDX3X immunoprecipitation and the IgG control immunoprecipitation (H=high; M=medium; L=low nucleotide size range).

4.4 DDX3X binds exons within translation factors mRNAs

The same bioinformatic approach used for DDX54 iCLIP was used to analyse DDX3X iCLIP data (Appendix Figure 1). This analysis showed DDX3X preferentially binds exons of protein coding RNAs (Figure 4.2 A,C) and it is preferentially found at 5′ and 3′ ends of transcripts with a predominant abundance towards the 5′ end (Figure 4.2B). Interestingly, in cell fractionation experiments, DDX3X was localized in both the cytoplasm and the nucleus (Figure 4.2D).

To validate binding of DDX3X to target RNAs, I performed CLIP-qPCR and analysed four DDX3X iCLIP target RNAs (*SOX4, TFF1, CALR, JUND*) and one RNA (*PUS1*) not bound by DDX3X (Figure 4.3). CLIP-qPCR was performed with either an anti-DDX3X antibody, antibodies against other DDX helicases (DDX1 and DDX24) or IgG control antibodies. Data show that DDX3X target RNAs were specifically and significantly enriched by DDX3X antibodies but not by the other antibodies (Figure 4.3). As expected, the control unbound RNA, *PUS1,* was unspecific in DDX3X CLIP-qPCR (Figure 4.3 bottom panel). These data indicate that the identified target RNAs are indeed specifically bound by DDX3X.

Next, DDX3X bound RNAs were subjected to a gene ontology analysis using the DAVID bioinformatics resource. This showed a striking predominance of genes encoding ribosomal components (Figure 4.4A-B). The human genome contains 80 ribosomal proteins genes of which 32 encode proteins of the small subunit and 48 encode proteins of the large subunit of ribosomes. Amongst the DDX3X target RNAs, I identified 42 (53% of total) ribosomal protein mRNAs, which constitute approximately 20% of all DDX3X bound RNAs (Figure 4.4C).

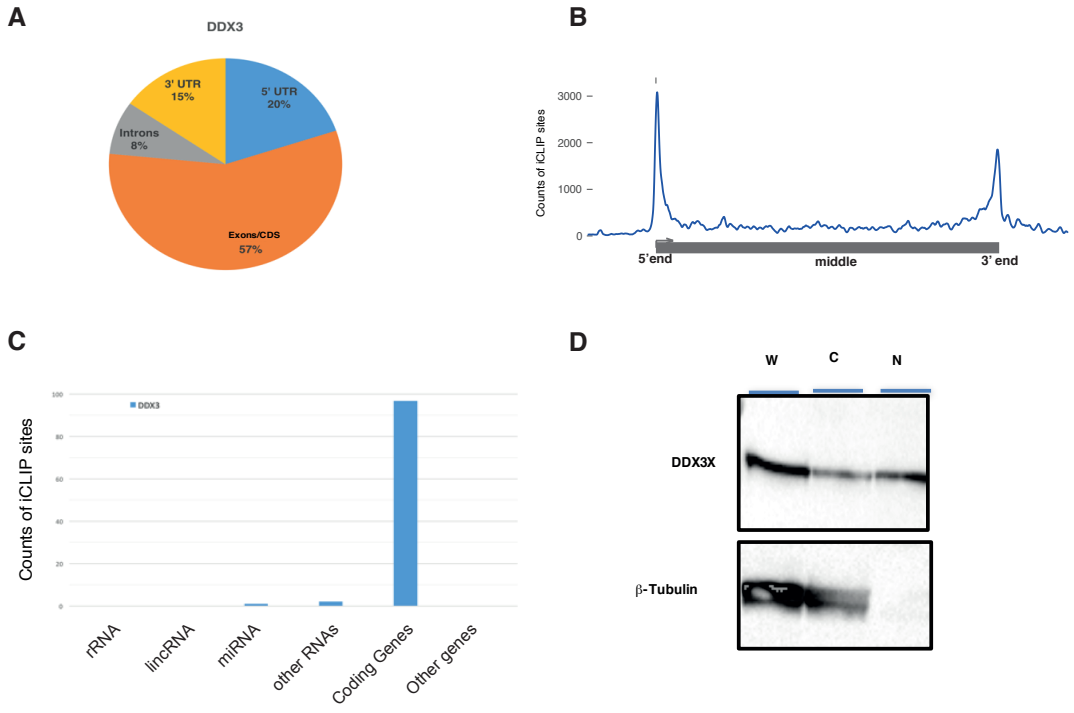


Figure 4.2. DDX3X iCLIP truncation sites localize on exonic regions and are enriched at 5' and 3' end of transcripts.
(A) Pie chart indicating location of iCLIP peaks in transcripts. The percentages of DDX3X iCLIP truncation sites found in the indicated categories are shown. Genomic categories of iCLIP peak location were searched for using PAVIS [PMID: 24008416] with default settings. (B) Density plot of iCLIP localisation on a scaled gene body (metaplot). The transcripts that have iCLIP peaks of DDX3X were scaled to the same length and also extended by 20% up- and down-stream. Then, iCLIP peaks for each mRNA were accumulated on the scaled plot to visualise the distribution of iCLIP truncation sites localisation. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis = counts of iCLIP sites; X axis = scaled gene body localization. (C) Bar graph of gene types with iCLIP peaks. Shown are the enrichments of DDX3X iCLIP truncation sites in each gene type category. The genomic categories of iCLIP peak location were searched for using PAVIS [PMID: 24008416] with default settings. Y axis = counts of iCLIP sites; X axis = gene type category. (D) DDX3X is localized in the nucleus and cytoplasm of MCF7 cells. Western blot analysis with antibodies against DDX3X and ß-tubulin, as indicated, in whole cell (W), cytoplasmic (C) and nuclear (N) fractions of MCF7 cells.

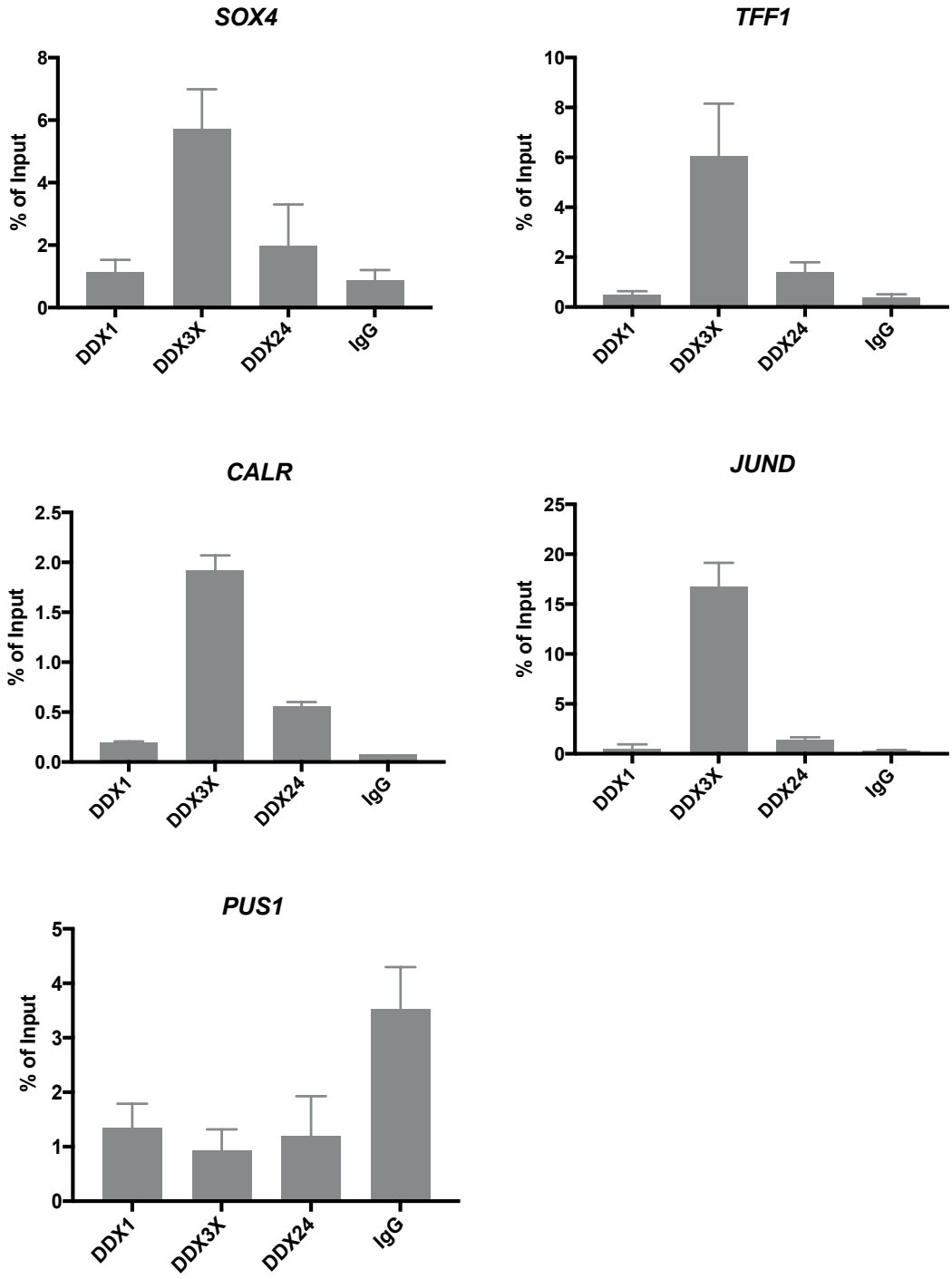
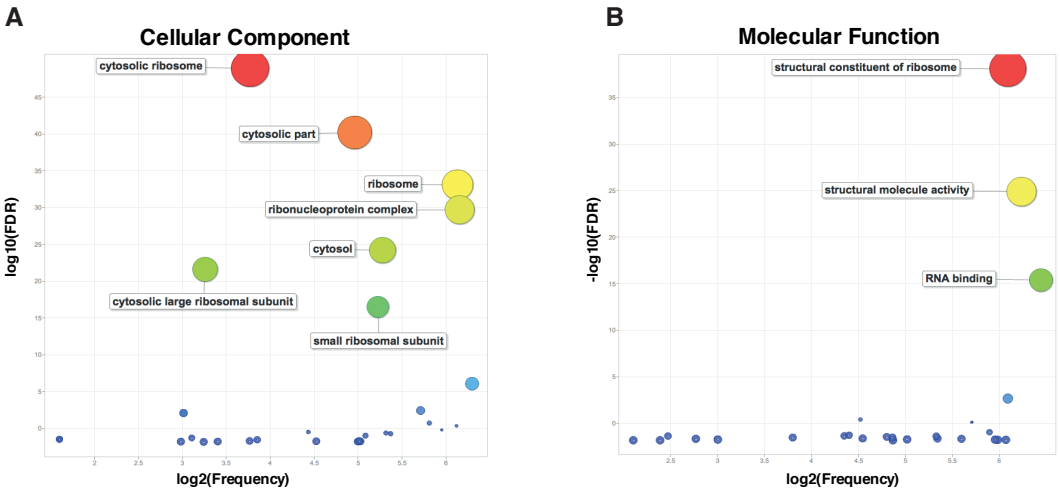


Figure 4.3. CLIP qPCR validation of DDX3X iCLIP targets.
CLIP-qPCR was used to validate binding of *SOX4, TFF1, CALR, JUND* mRNAs to DDX3X. DDX1, DDX24 and IgG were included as negative controls, as was the non-bound *PUS1* mRNA. cDNA obtained from each immunoprecipitation was run in qPCR to determine enrichment of the indicated targets. Results represent the average of three biological replicates. Error bars indicate standard deviation.



C

ACTB	COX7A2	HIST1H2AG	NQO1	RPL19	SLC1A5
ACTG1	COX7C	HIST1H2BD	NR2F2	RPL27A	SLC7A2
ADNP	CSDE1	HN1	NUTF2	RPL28	SLC7A5
AGBL5	CTSD	HNRNPA2B1	OPA1	RPL30	SLC9A3R1
AHNAK	CTTN	HNRNPF	PABPC1	RPL31	SLC25A3
AHSA1	CXXC5	HSP90AA1	PARD6B	RPL34	SLC25A37
ALDOA	DAD1	HSP90AB1	PARK7	RPL35	SLC38A1
ANAPC5	DBI	HSPA5	PCBP2	RPL37	SLC39A6
APH1A	DDB1	HSPA8	PDXK	RPL37A	SLMO2
ARPC2	DDX5	HSPB1	PFKL	RPL38	SNHG7
ATF4	DKK1	IGF1R	PGK1	RPLP0	SNRPC
ATP1A1	DPP9	IGSF3	PKM	RPLP1	SNRPF
ATP5A1	DSP	JUND	PLEC	RPLP2	SOD1
ATP5B	EEF1A1	JUP	PMEPA1	RPS2	SOX4
ATP5E	EEF1B2	KCNJ3	PPDPF	RPS3	SPINT2
ATP5G3	EEF2	KCTD2	PPIA	RPS3A	SPTSSB
ATP6V0E1	EIF4B	KDELR1	PPP2R1A	RPS5	SRP14
AZIN1	EIF4G2	KLF4	PPP4C	RPS6	SRRM2
B2M	EMP2	KRT8	PRDX1	RPS8	SRSF6
BCAP31	ENO1	KRT18	PRRC2A	RPS12	SSR2
BMP7	EPCAM	KRT19	PRSS23	RPS13	SSR4
BRK1	ERBB3	KRT80	PSAP	RPS15A	SULF2
BTF3	EVL	LAPTM4B	PSMA4	RPS16	SYNGR2
CACNG4	FAM83H	LARP4	PSMD4	RPS18	TCEB2
CALR	FAM83H-AS1	LGALS1	PTMA	RPS19	TFF1
CCND1	FAU	LPCAT1	RAC1	RPS20	TFF3
CCT5	FLNA	LRP10	RAD23B	RPS21	THEM6
CD47	FOXA1	LSR	RAN	RPS24	TMBIM6
CD59	FTH1	LY6E	RHBDD2	RPS25	TMED10
CD63	FTL	MAGED1	RNF168	RPS26	TMEM64
CDH1	FXYD3	MAZ	RPL3	RPS27	TMEM222
CDKN1A	GAPDH	MCL1	RPL4	RPS27A	TMSB4X
CELSR2	GATA3	MDH2	RPL5	RPS27L	TPT1
CERS2	GFRA1	MDM2	RPL6	RPS28	TUBB
CHCHD5	GNAS	METRN	RPL7A	RPS29	TXN
CHPF2	GNB2L1	MYL12B	RPL8	RPSA	TXNIP
CLDN4	GPRC5A	NCOA3	RPL10	S100A10	TXNL4A
CLDN7	GPX4	NDUFA4	RPL10A	SCAMP2	USF1
CLTC	GREB1	NDUFB9	RPL11	SCD	USMG5
CNIH4	H3F3B	NEAT1	RPL12	SDC1	VMP1
CNOT6L	HDLBP	NEDD4L	RPL13	SEC61B	XBP1
COX4I1	HINT1	NME4	RPL14	SEMA3C	YWHAZ
COX5B	HIPK1	NOP10	RPL18	SEPT9	ZNF217
COX6C	HIST1H2AC	NPM1	RPL18A	SIVA1	ZNF767P

Figure 4.4. DDX3X iCLIP target RNA gene ontology analysis shows enrichment of ribosomal components.
DAVID gene ontology search of cellular component (A) and molecular function (B) enriched terms. The GO enrichment of DDX3X iCLIP bound transcripts was obtained using the DAVID functional annotation tool with default settings [PMID: 19131956]. P-values were calculated using the "default" method of DAVID.

(A-B) Scatter plots were generated with REVIGO [PMID: 21789182] to visualise the GO term analysis result for cellular components (A) and molecular function (B). The scatter plots show GO term cluster representatives in two-dimensional space. X-axis indicates the frequency of the GO term in the underlying GO analysis database (i.e. bubbles that are closer to the end of x-axis (lager values) are more general terms). Y-axis and bubble colour/size indicates the log transformed p-value of GO terms. (C) Table shows all RNA targets of DDX3X iCLIP and highlights ribosomal proteins.

As with the RNAs validated above (Figure 4.3), I validated binding of DDX3X to specific ribosomal protein mRNAs (*RPS2, RPS3* and *RPS6*) by performing CLIP-qPCR (Figure 4.5A). Each of these mRNAs preferentially bound to DDX3X. Importantly, the specificity of the anti-DDX3X antibody was corroborated by performing CLIP-qPCR in MCF7 cells that had been depleted of DDX3X using shRNA. In these experiments, precipitation of each ribosomal protein mRNA was dramatically reduced upon loss of DDX3X (Figure 4.5B). Efficient knock down of DDX3X was confirmed by western blotting with the anti-DDX3X antibody (Figure 4.5C).

To investigate the significance of DDX3X to the synthesis of ribosomal proteins, I analysed the levels of RPS2, RPS3 and RPS6 proteins, via western blotting with appropriate antibodies upon knock down of DDX3X (via shRNA and siRNAs) in MCF7 cells (Figure 4.6A). RPS2 and RPS6 were depleted upon loss of DDX3X, with the highest effect observed on RPS2 protein level, whilst RPS3 and the loading control β-tubulin remained unchanged (Figure 4.6A). At the mRNA level, *RPS2, RPS3* and *RPS6* were slightly reduced in both transient siRNA-mediated and stable shRNA-mediated DDX3X knocked down MCF7 cells (Figure 4.6B). Overall, these results suggest that binding of DDX3X to ribosomal protein mRNAs may be required for their efficient translation and for stability of *RPS2, RPS3* and *RPS6* mRNAs.

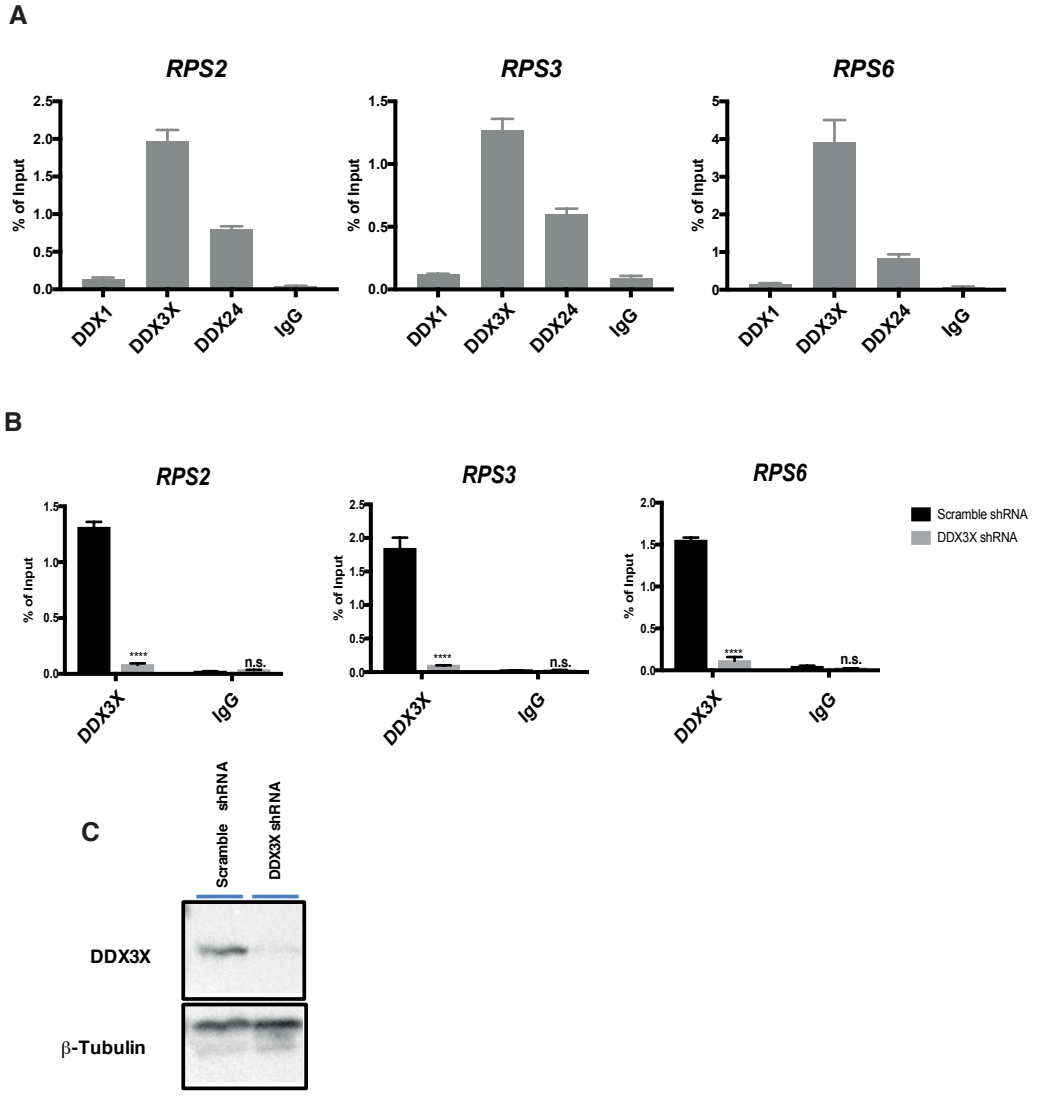


Figure 4.5. CLIP-qPCR validation of DDX3X binding to ribosomal protein mRNAs.

(A) Clip-qPCR was performed to validate binding of *RPS2, RPS3*, and *RPS6* mRNAs to DDX3X. DDX1, DDX24 and IgG were included as negative controls. (B) DDX3X and IgG CLIP-qPCR were performed in MCF7 cells transduced with scrambled or DDX3X targeting shRNA. (A-B) cDNA obtained from each immunoprecipitation was analysed by qPCR to test for enrichment of the indicated targets. Results represent the average of three biological replicates. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001; **** ≤ 0.0001). (C) Western blot analysis shows the level of DDX3X and ß-tubulin in whole lysate of MCF7 cells transduced with scrambled or DDX3X targeting shRNA, used in B.

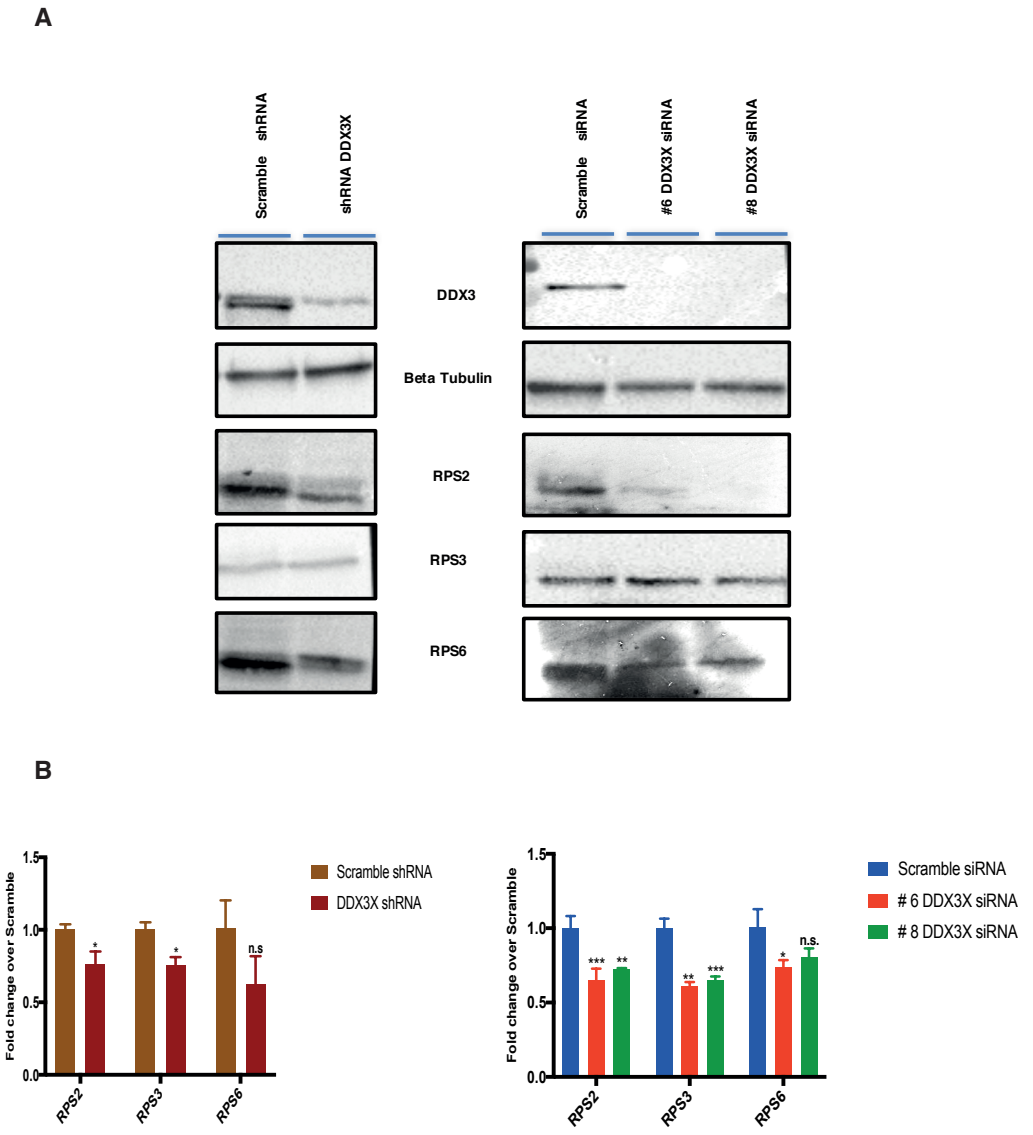


Figure 4.6. DDX3X knock down affects ribosomal proteins mRNA and protein levels.

(A) Western blot analysis shows the level of RPS2, RPS3, RPS6, DDX3X and ß-tubulin proteins in whole lysate from MCF7 cells transduced with either scrambled shRNA or DDX3X targeting shRNA (left panel) and in cell lysate from MCF7 cells transfected with either scrambled siRNA or two different DDX3X targeting siRNAs (#6 or #8) (right panel). (B) *RPS2, RPS3 and RPS6* mRNA levels in MCF7 treated as in A, measured by qPCR. Results represent the average of three biological replicates (Section 2.5.3). Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble controls (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001**** ≤ 0.0001).

4.5 Does DDX3X regulate ribosomal protein mRNAs in mTOR signalling?

In order to find pathways in which DDX3X may function, I ran additional gene ontology analysis on pathways enrichment. As expected, my search again indicated that most DDX3X target RNAs fall into translational regulation pathways. Particularly, mTOR (mechanistic target of rapamycin) signalling was one of the top terms (Figure 4.7). Indeed, most of the genes listed in the mTOR signalling category were ribosomal proteins, amongst which were the validated DDX3X binding targets *RPS2, RPS3* and *RPS6*.

As previously described, mTORC1 is a known regulator of translation and is required mainly for translation of translation factors mRNAs, many of which encode ribosomal proteins (Section 1.1.3) [77]. Furthermore, DDX3X binds mRNA 5′-caps in an mTOR-dependent manner. DDX3X association to the 5′-cap complex is increased upon insulin dependent mTOR signalling activation and reduced upon mTOR signalling chemical inhibition [96]. These reports, together with the results from my bioinformatic analysis, prompted me to investigate whether DDX3X binding to ribosomal protein mRNAs could affect their protein levels in an mTOR-dependent manner. I speculated that if mTOR regulates binding of DDX3X to its target mRNAs, then inhibition of the mTOR pathway (by rapamycin treatment) should affect DDX3X binding to its targets. To test this hypothesis, I subjected MCF7 cells to a time-course of rapamycin treatment. I first performed a short-term time-course (0-6 h) using 100 nM rapamycin. This showed no changes in either the protein or mRNA levels of *RPS2, RPS3* and *RPS6*, although the mTOR pathway was efficiently inhibited as demonstrated by loss of a signalling marker, namely RPS6 phosphorylation (Figure 4.8A-B). DDX3X binding to ribosomal proteins mRNAs was also unaffected over this short-term time-course of rapamycin treatment (Figure 4.8C).

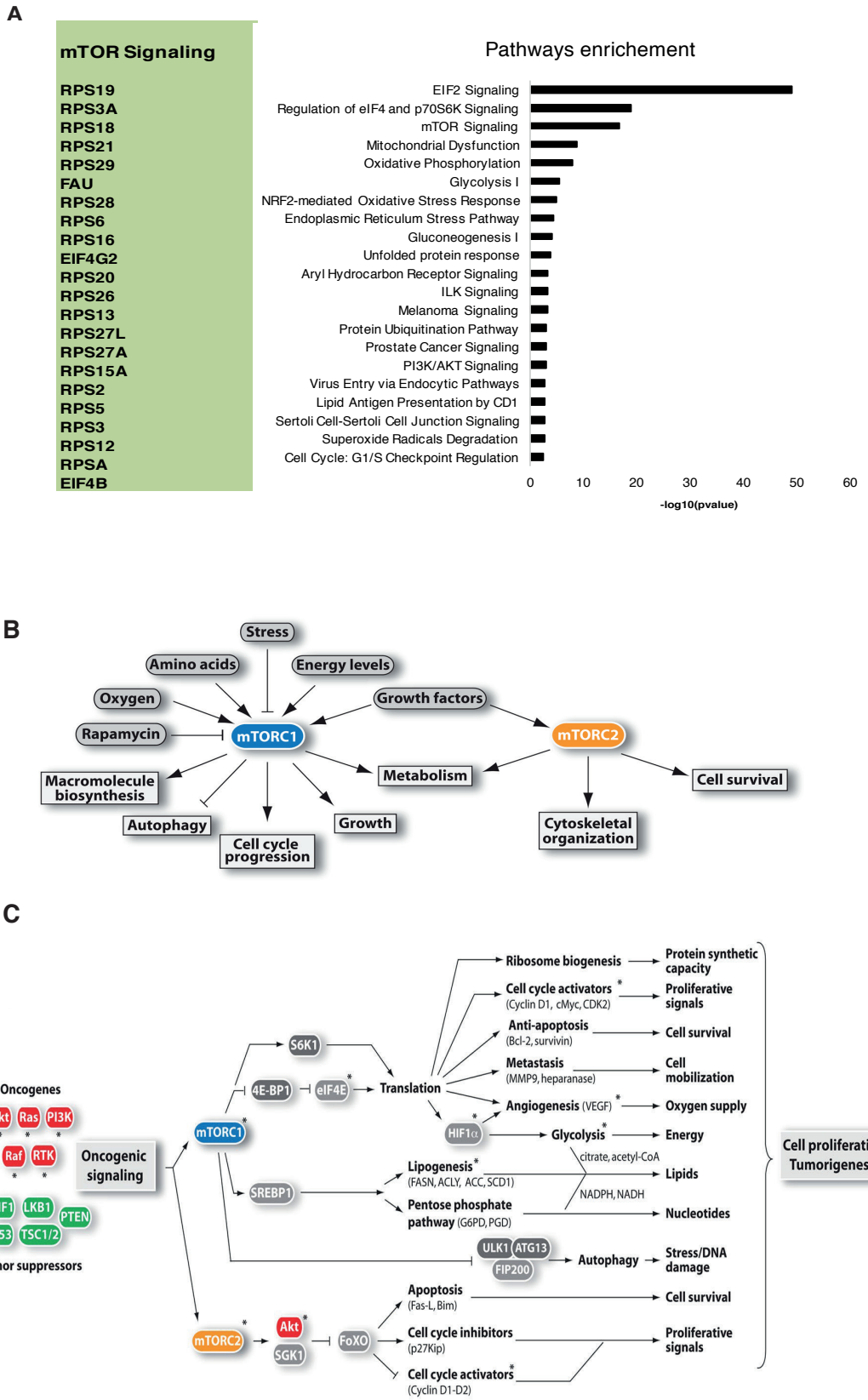


Figure 4.7. Pathways enrichment of DDX3X iCLIP targets: mTOR signalling. (A). Bar graph showing pathways term enrichment generated with IPA tool (IPA Ingenuity® Systems) GO search of DDX3X iCLIP targets. P-values were

calculated using default settings of Ingenuity. The X-axis shows GO terms related to upstream regulators. Y-axis shows –log10 p-values of each GO term. The table shows the components of the mTOR term: mostly ribosomal proteins. (B) The mTOR kinase is the core component of two distinct protein complexes termed mTORC1 and mTORC2. mTORC1 responds to amino acids, stress, oxygen, energy, and growth factors and is acutely sensitive to rapamycin. It promotes cell growth by inducing and inhibiting anabolic and catabolic processes, respectively, and also drives cell cycle progression. mTORC2 responds to growth factors and regulates cell survival and metabolism, as well as the cytoskeleton. mTORC2 is insensitive to acute rapamycin treatment but chronic exposure to the drug can disrupt its structure. (C) mTOR signaling promotes tumorigenesis. Oncogenes (red) or tumor suppressors (green) implicated in the control of mTOR signaling are indicated. Adapted from [77]

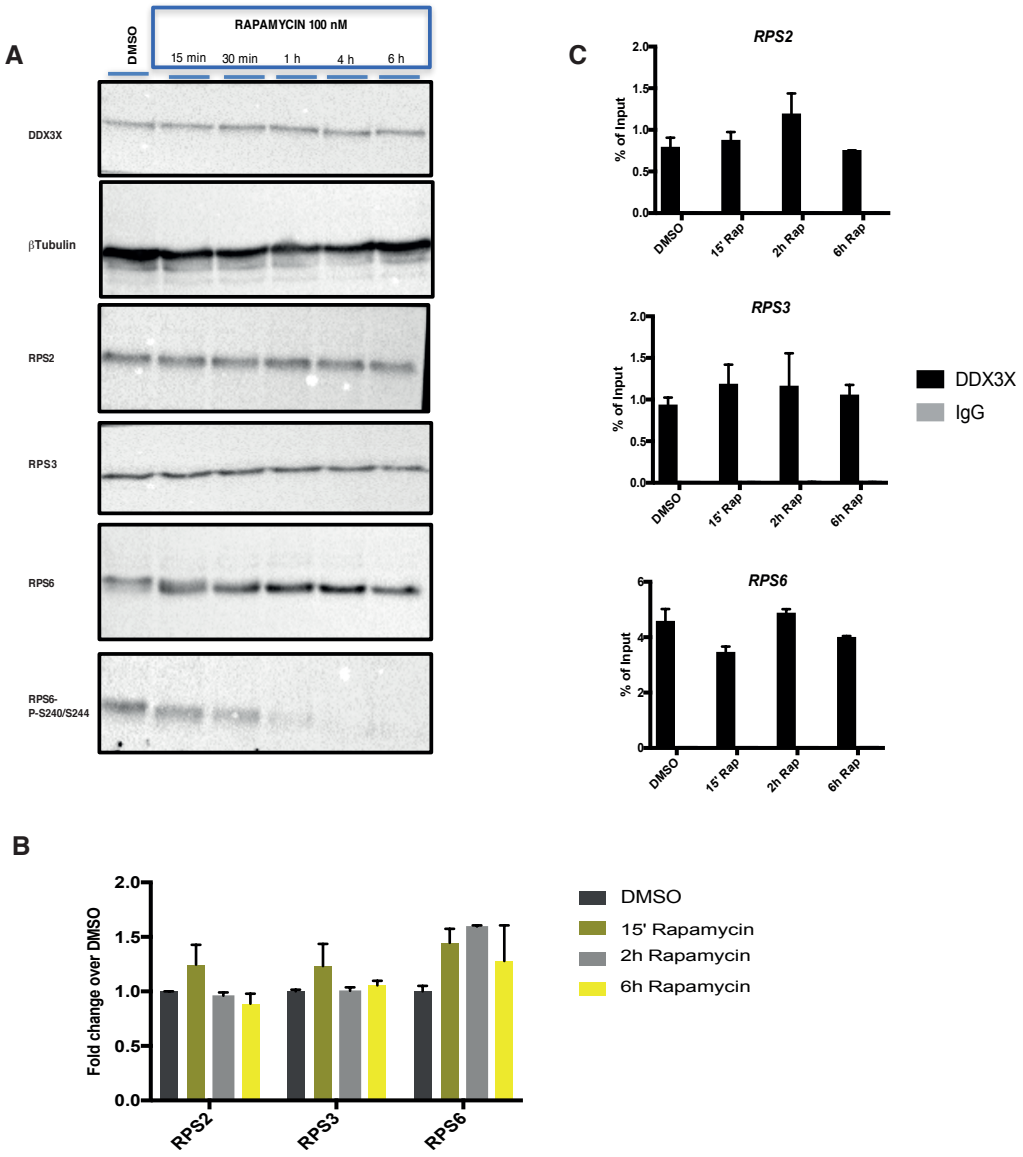


Figure 4.8. Effects of rapamycin inhibition of mTOR pathway at early time points on RPS2, RPS3 and RPS6 expression and DDX3X binding of their respective mRNAs.

(A) Western blot shows the level of RPS2, RPS3 RPS6, DDX3X and ß-tubulin at the indicated time points upon treatment with 100 nM rapamycin or DMSO control. RPS6 phosphorylation (at residues S240/S244) reduction is used as a marker of mTOR activity inhibition. (B) mRNA levels of *RPS2, RPS3* and *RPS6* in MCF7 cells treated with DMSO or with 100 nM rapamycin at the indicated time-points, measured by qPCR. Results represent the average of three biological replicates (Section 2.5.3). P-values were n.s. (C) DDX3X and IgG CLIP-qPCR performed as previously described to test the binding of RPS2, RPS3 and RPS6 mRNAs in MCF7 cells treated with DMSO or with 100 nM rapamycin at the indicated time points. cDNA obtained from each immunoprecipitation was analysed by qPCR to determine enrichment of the indicated targets. Results represent the average of three biological replicates. Error bars indicate standard deviation. P-values were n.s.

Next, I exposed MCF7 cells to rapamycin for 24 hours, and compared the effects of this longer treatment on the expression of RPS2, RPS3 and RPS6 proteins in MCF7 cells in which DDX3X was knocked down or normally expressed (Figure 4.9A). The intensity of the bands in these western blots were then quantified using ImageLab (Figure 4.9B). Strikingly, after 24 h of rapamycin treatment, the level of DDX3X itself was reduced by about 70% compared to the level in control DMSO treated cells. Interestingly, knock down of DDX3X under DMSO control conditions was as effective as the rapamycin treatment in reducing the level of DDX3X protein. However, rapamycin treatment caused a further reduction of the DDX3X protein level in DDX3X knock down cells. A similar trend was observed for RPS2 protein. Conversely, RPS3 protein levels were not affected by DDX3X knock down but were decreased upon 24 h rapamycin treatment. RPS6 protein levels were insensitive to rapamycin treatment but were reduced by DDX3X knock down, equally in either DMSO or rapamycin treated cells. Interestingly, mRNA levels of *RPS2, RPS3* and *RPS6* were reduced in DDX3X knockdown cells and went slightly up in the presence of rapamycin inhibition in cell lines expressing DDX3X as well as in those with DDX3X knock down (Figure 4.9C).

CLIP-qPCR performed in the same conditions showed that after 24 h of rapamycin treatment, binding of DDX3X to *RPS2* and *RPS6* mRNAs was almost unchanged, whilst binding to *RPS3* mRNA was slightly increased, compared to the binding detected in the DMSO controls (Figure 4.9D). However, DDX3X protein level is strongly decreased after 24h rapamycin treatment (Figure 4.9B). Therefore, assuming an equal efficiency of DDX3X immunoprecipitation for all samples, one could speculate that the residual DDX3X protein expressed after 24h of rapamycin treatment is mostly engaged in binding of its target mRNAs.

These results show that DDX3X binding to *RPS2, RPS3* and *RPS6* mRNAs is maintained upon inhibition of the mTOR pathway with rapamycin. Moreover, the levels of RPS2, RPS3 and RPS6 proteins were differently affected by both rapamycin treatment and DDX3X knock down: RPS2 is

strongly decreased by DDX3X knock down as well as by rapamycin treatment; RPS3 is insensitive to DDX3X knock down but decreased upon rapamycin treatment; RPS6 is insensitive to rapamycin and decreased upon DDX3X knockdown (although showing less sensitivity than RPS2). These observations indicate the complexity of the hypothesized connection between DDX3X and the mTOR pathway.

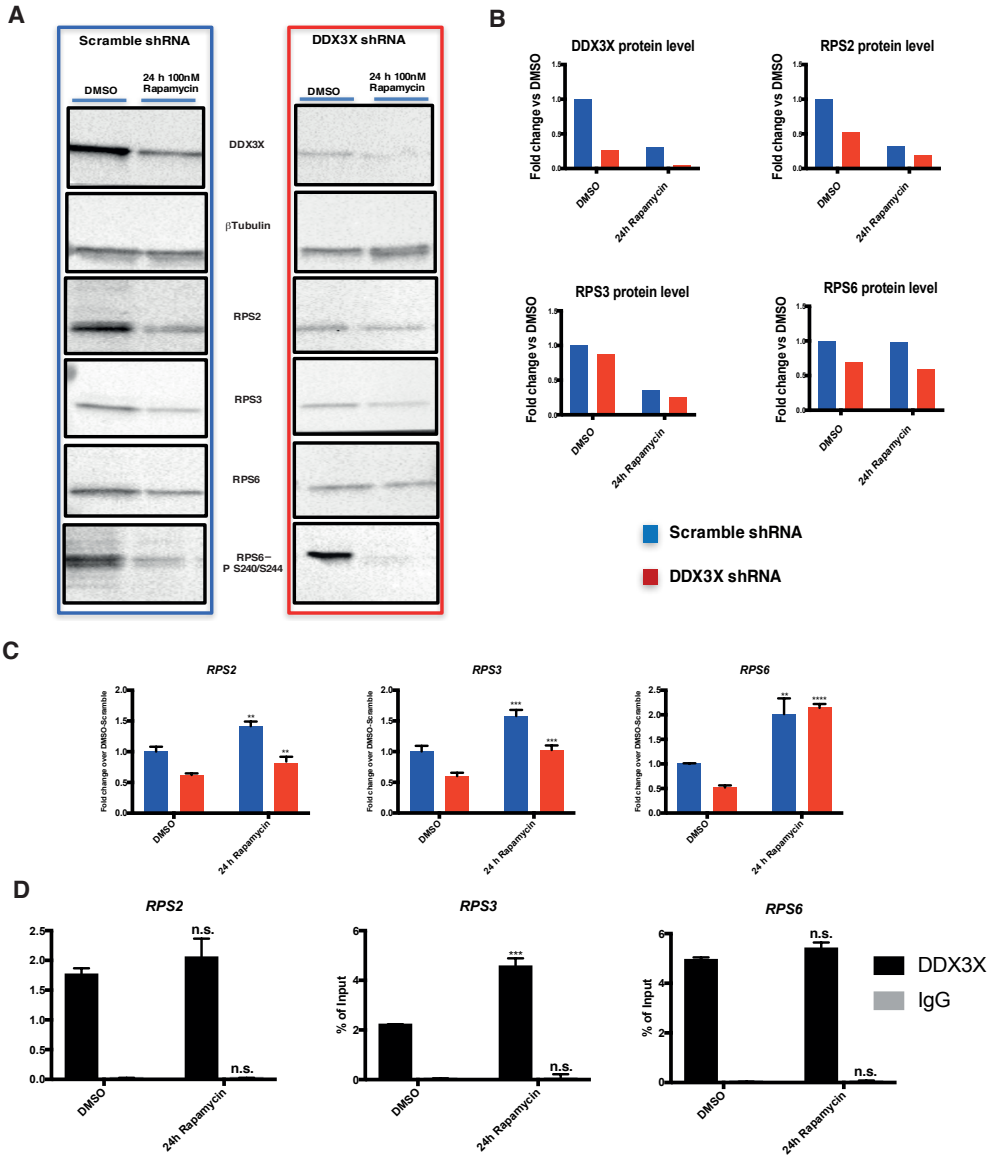


Figure 4.9. Effects of 24h rapamycin inhibition of the mTOR pathway on RPS2, RPS3 and RPS6 expression and DDX3X binding of their respective mRNAs.
(A) Western blot shows the levels of RPS2, RPS3, RPS6, DDX3X and ß-tubulin proteins in MCF7 cells transduced with either scrambled shRNA or DDX3X targeting shRNA and treated with DMSO or 100 nM rapamycin for 24 h.

Reduction of RPS6 phosphorylation (residues S240/S244) was used as a marker of mTOR activity inhibition. (B) protein levels from A were quantified with ImageLab. Values were normalized to ß-tubulin levels and expressed as fold over DMSO control.
(C) mRNA levels of RPS2, RPS3 and RPS6 tested under the same conditions as in A, measured by qPCR. (D) DDX3X and IgG CLIP-qPCR performed as previously described to detect binding of RPS2, RPS3 and RPS6 mRNAs in MCF7 cells treated with DMSO or 100 nM rapamycin for 24h. cDNA obtained from each immunoprecipitation was analysed by qPCR to determine enrichment of the indicated targets. (C-D) Results represent the average of three biological replicates (Section 2.5.3). Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to DMSO (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

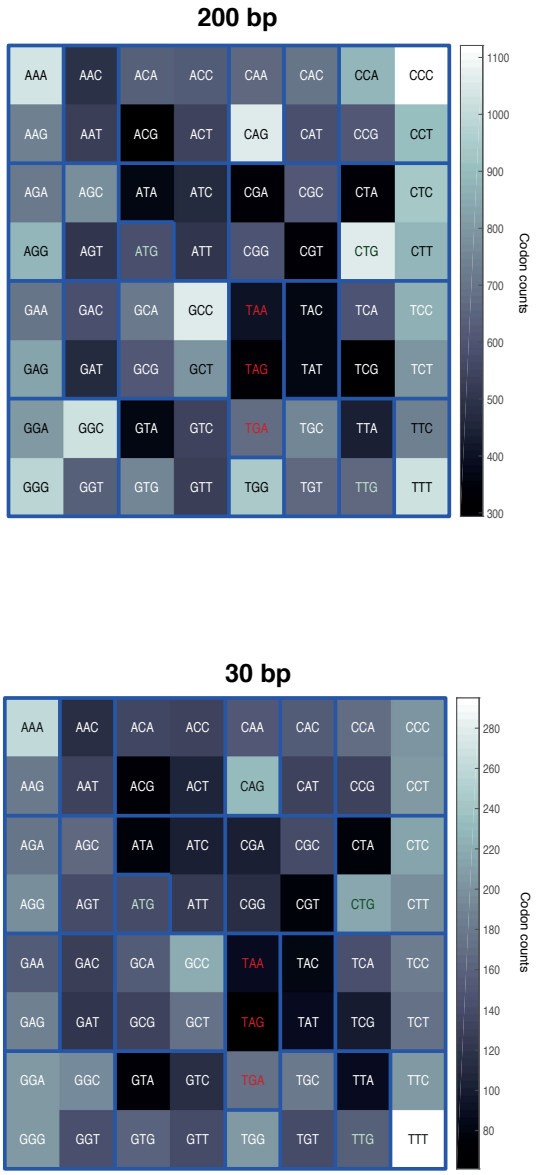
4.6 DDX3X binds specific sequences within its target RNAs.

To further investigate binding specificity of DDX3X, I analysed RNA sequences surrounding DDX3X binding sites and observed enrichment of CG and CT (i.e. CU in RNA) containing triplets in both 200 base-pair and 30 base-pair windows around the truncation sites (Figure 4.10A). To ask whether these triplets could be part of a more enriched sequence binding motif, I ran a motif search and found ten subgroups of related 8-mers to which most DDX3X targets could be assigned (Figure 4.10B). All enriched motifs were localized towards either the 5′- or the 3′-end of transcripts, consistent with the truncation site localization previously observed (Figure 4.2B and Figure 4.10C), which suggests most target RNAs may be bound via a discrete motif.

Amongst the most enriched motifs were subgroups (Figure 4.11) showing high similarity to known binding sites of RNA binding proteins involved in different molecular processes (data not shown). Of most relevance, the motif called SG2 was found to possess strong sequence homology with the binding motif of SR splicing factors 1 and 2 (SRSF1, SRSF2), suggesting that DDX3X may be involved in alternative splicing, at least of RNAs bearing this motif (Figure 4.11B: SG2).

My analyses also revealed that ribosomal protein RNAs generally bear more than one motif spread throughout their mRNA sequence and, although belonging to different phylogenetic subgroups, most of the motifs present within these RNAs share a CT (i.e. CU in RNA) enrichment within the 8-mer sequence, as shown in the examples in Figure 4.11A-D. Interestingly, the three validated ribosomal proteins mRNAs (*RPS2, RPS3* and *RPS6*) bear different binding motifs within their sequences (Figure 4.12).

A


200 bp
AAA AAC ACA ACC CAA CAC CCA CCC
AAG AAT ACG ACT CAG CAT CCG CCT
AGA AGC ATA ATC CGA CGC CTA CTC
AGG AGT ATG ATT CGG CGT CTG CTT
GAA GAC GCA GCC TAA TAC TCA TCC
GAG GAT GCG GCT TAG TAT TCG TCT
GGA GGC GTA GTC TGA TGC TTA TTC
GGG GGT GTG GTT TGG TGT TTG TTT
1100
1000
900
800
700
600
500
400
300
Codon counts
30 bp
AAA AAC ACA ACC CAA CAC CCA CCC
AAG AAT ACG ACT CAG CAT CCG CCT
AGA AGC ATA ATC CGA CGC CTA CTC
AGG AGT ATG ATT CGG CGT CTG CTT
GAA GAC GCA GCC TAA TAC TCA TCC
GAG GAT GCG GCT TAG TAT TCG TCT
GGA GGC GTA GTC TGA TGC TTA TTC
GGG GGT GTG GTT TGG TGT TTG TTT
280
260
240
220
200
180
160
140
120
100
80
Codon counts

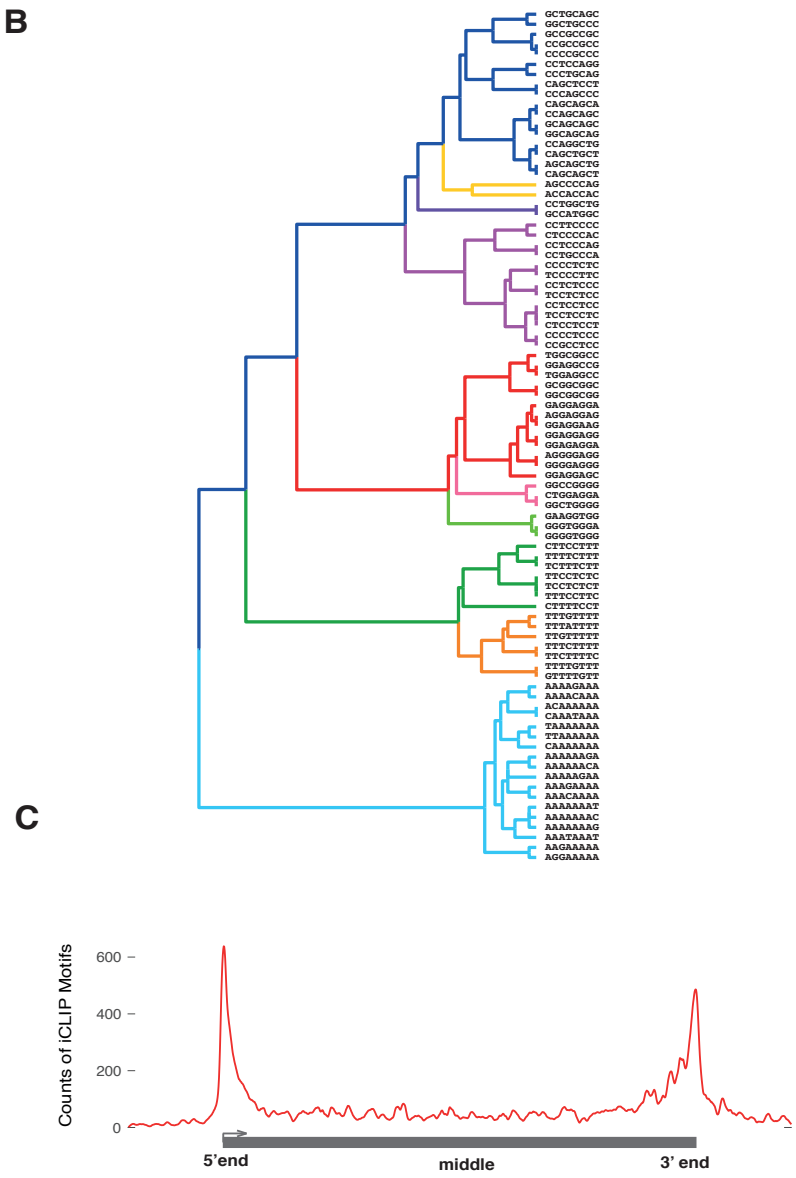


Figure 4.10. Motif search on DDX3X iCLIP binding sites.
(A) Heatmap of triplets usage in 200 base-pair and 30 bp windows centred by DDX3X iCLIP peak. Shown is the enrichment of triplets found in a 200 bp (left) and 30 bp (right) extended window centred on iCLIP peaks. MATLAB *codoncount* was used to generate the heatmap (Section 2.13.5). The colour gradient bar indicates counts of triplets found in the considered window and associated hue. (B) Phylogenetic tree of highly enriched 8-mer motifs in 200 base-pair windows centred by DDX3X iCLIP peak. 8-mers enriched in an extended +/-100bp centred by iCLIP were found as described (Section 2.13.5). To identify consensus motifs, enriched 8-mers were phylogenetically clustered into 10 groups. MATLAB function *seqlinkage* was used to construct the phylogenetic tree shown (Section 2.13.5). (C) Density plot of highly enriched motifs localised on scaled gene body. Shown is the distribution of the motif localisation. The transcripts that have the enriched motifs in the +/-100bp extended windows centred by DDX3X iCLIP

peaks were scaled to the same length and also extended by 20% up- and down-stream. Then, the enriched motifs for each gene were accumulated on the scaled gene to visualise the distribution of DDX3X binding motif localisation. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis = counts of iCLIP Motifs; X axis = scaled gene body localization.

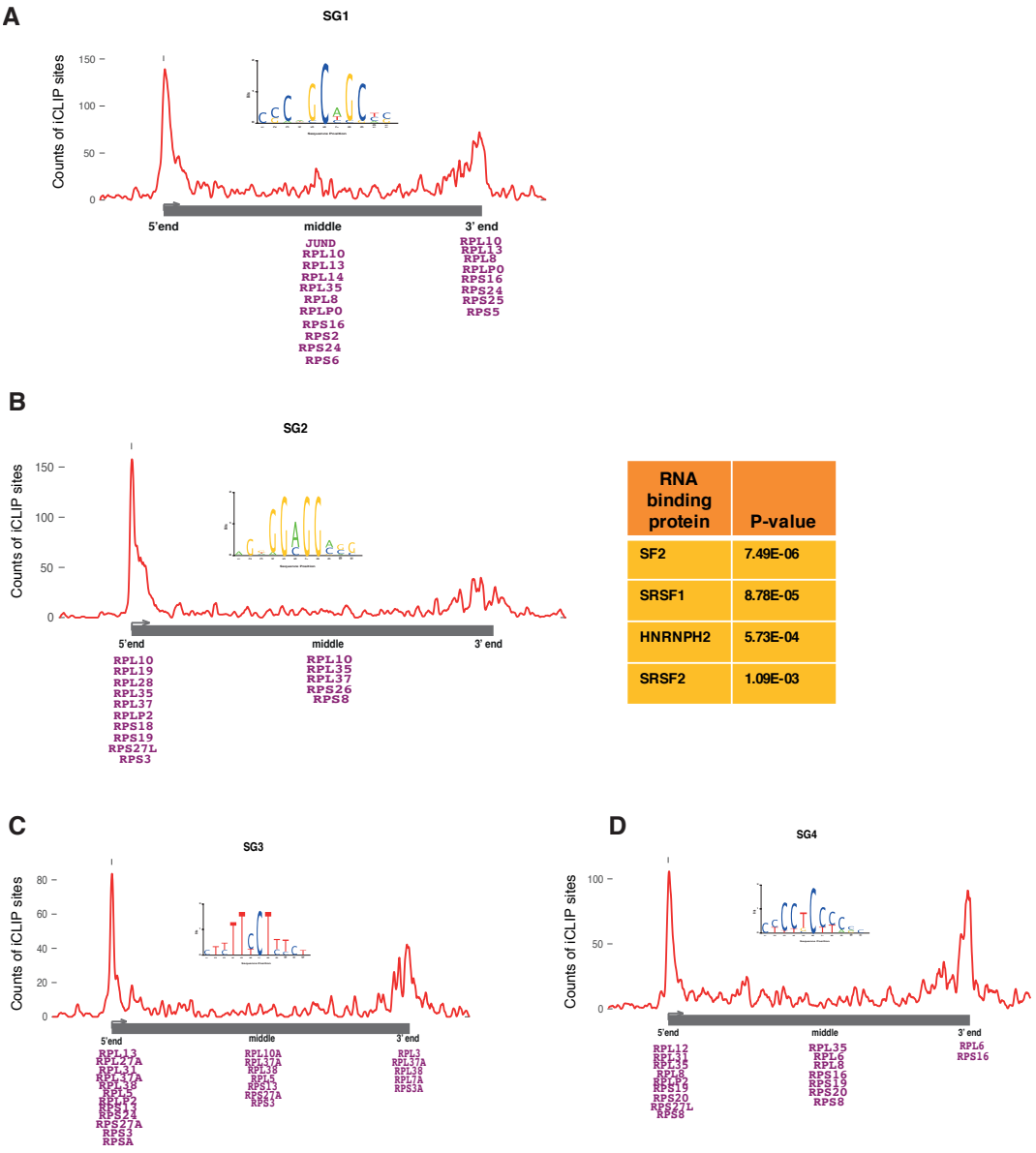


Figure 4.11. Enriched motifs at DDX3X binding sites.
(A-D) Shown are 4 examples of enriched motifs at DDX3X binding sites together with a density plot of the motifs localisation within a scaled gene body. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis indicates the frequency (actual count) of the motif; X axis = scaled gene body localization. Consensus motifs for the clustered 8-mer(s) in each phylogenetic group were found with MATLAB function *seqlogo* (Section 2.13.5). Each density plot represents the localization of the clustered motifs of four different phylogenetic groups considered (SG1, SG2, SG3, SG4). The sequence logo of the phylogenetic group consensus motif is shown (Y axis = Bits; X axis =sequence position). Ribosomal proteins and *JUND* bearing the motif are listed under the corresponding plot at the scaled position where the motif was found in their sequence. (B) The motif is matched to known RNA binding motifs. Table shows the known RNA binding proteins that recognize a similar sequence motif. P-values represent the significance of homology.

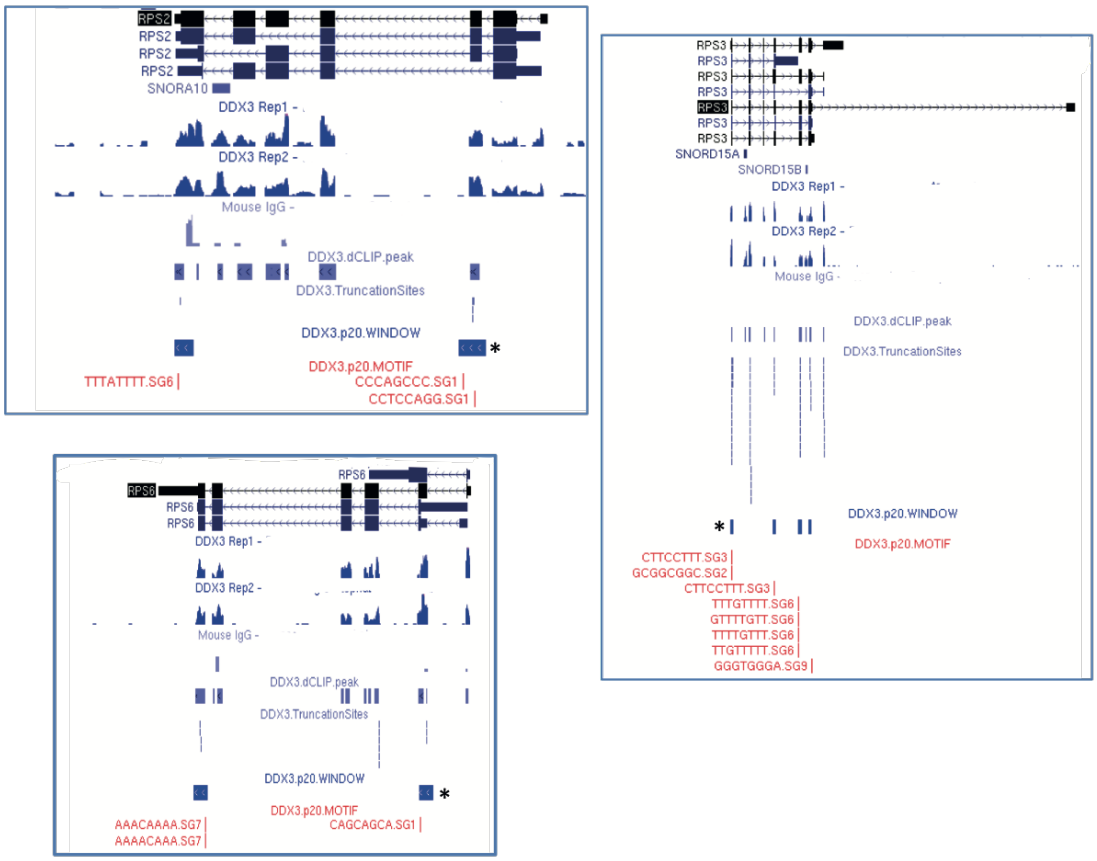


Figure 4.12. DDX3X binding motif localization within *RPS2, RPS3* and *RPS6* mRNAs.
Visualization of the USCS custom tracks for DDX3X iCLIP truncation sites and motif localization within *RPS2*, *RPS3*, and *RPS6* mRNA sequences.

4.7 DDX3X binds *hJUND* mRNA at a specific binding motif.

The *JUND* transcript was shown to require binding of the DEAH BOX RNA helicase DHX9 for efficient translation **[247]**. In my CLIP-qPCR validations, the *JUND* transcript was one of the most enriched and highly reproducible DDX3X target RNAs detected (Figure 4.3). Furthermore, the specificity of DDX3X binding to *JUND* RNA was confirmed by CLIP-qPCR experiments in DDX3X knock down cells (Figure 4.13A).

Interestingly, both *JUND* mRNA and protein levels showed significant depletion upon DDX3X knock down in both DDX3X shRNA transduced and DDX3X siRNA transfected MCF7 cells (Figure 4.13B-C). Moreover, my motif search analysis identified a CT- (i.e. CU in RNA) rich putative DDX3X binding motif within the *JUND* transcript (Figure 4.14A marked with *****). These observations, together with published reports, provided the rationale to validate the motif search results within the sequence of *JUND* mRNA. For this purpose, I performed electrophoretic mobility shift assays (EMSA) using *in vitro* synthesized RNA oligonucleotides. I designed a wild type RNA oligonucleotide that included 52 nucleotides surrounding the putative DDX3X binding motif site within *JUND* mRNA. I also designed a mutant RNA oligonucleotide mapping to the same region but harbouring a three-nucleotide substitution within the motif 8-mer (Figure 4.14B). The oligonucleotides were then 5'-end labelled with ^{32}P and incubated with MCF7 whole cell lysate prepared from cells that had been transduced with either scrambled shRNA or DDX3X shRNA. The oligonucleotides and bound proteins were then resolved in a polyacrylamide non-denaturing gel via standard EMSA procedures (Section 2.7). The autoradiograph of the resulting gel showed free probe at the bottom of the gel with a number of 'shifted' higher molecular weight bands above. The bands with lower mobility than free probe represent protein-RNA complexes.

The band with an intermediate mobility (marked with *) was only observed with the WT probe and not the mutant probe (Figure 4.14C).

Importantly, this band was completely lost when MCF7 cell extracts depleted of DDX3X were used (Figure 4.14C, lane IV). This suggests that DDX3X is essential for the formation of this protein-RNA complex and this signal represents a DDX3X-RNA complex.

The band with the lowest mobility (marked with ^) was present with both the wild type probe and mutant probe. Interestingly, this band ran slightly faster when the wild type probe was used in binding reactions with lysate from DDX3X knock down MCF7 cells (Figure 4.14C – WT *JUND* probe box, lane IV). This suggests that DDX3X is part of a larger protein-RNA complex and when it is depleted, other components of the complex can still bind to the probe.

When the mutant *JUND* probe was used in binding reactions with lysate from DDX3X-expressing MCF7 cells, the mobility of this band was intermediate compared to the mobility shown when using the wild type probe (Figure 4.14C - MUTANT *JUND* probe box, lanes II) with lysates expressing or depleted of DDX3X (Figure 4.14C - WT *JUND* probe box, lanes II and IV). When the mutant *JUND* probe was used in binding reactions with lysate from DDX3X knock down MCF7 cells the intensity of this band was reduced (Figure 4.14C - MUTANT *JUND* probe box, lanes IV) but not its mobility.

Overall, these results suggest that DDX3X binds directly to the identified motif within *JUND* mRNA (* band) and that a multicomponent complex containing DDX3X specifically recognizes the motif within the *JUND* sequence (^ band). In addition, other factors within the complex appear capable of binding to the probe, outside of the specific DDX3X binding motif, independently of DDX3X. Finally, it is likely that DDX3X also binds to other regions of the probe and/or interacts with other factors of the complex, to mediate co-operative binding of the rest of the complex.

To test the specificity of DDX3X binding to the *JUND* oligonucleotides, I performed an antibody super-shift experiment by adding an anti-DDX3X antibody to the binding reactions (performed as previously described).

Unfortunately, no change in band mobility was observed with either wild type or mutant *JUND* probes (Figure 4.14C- lanes III).

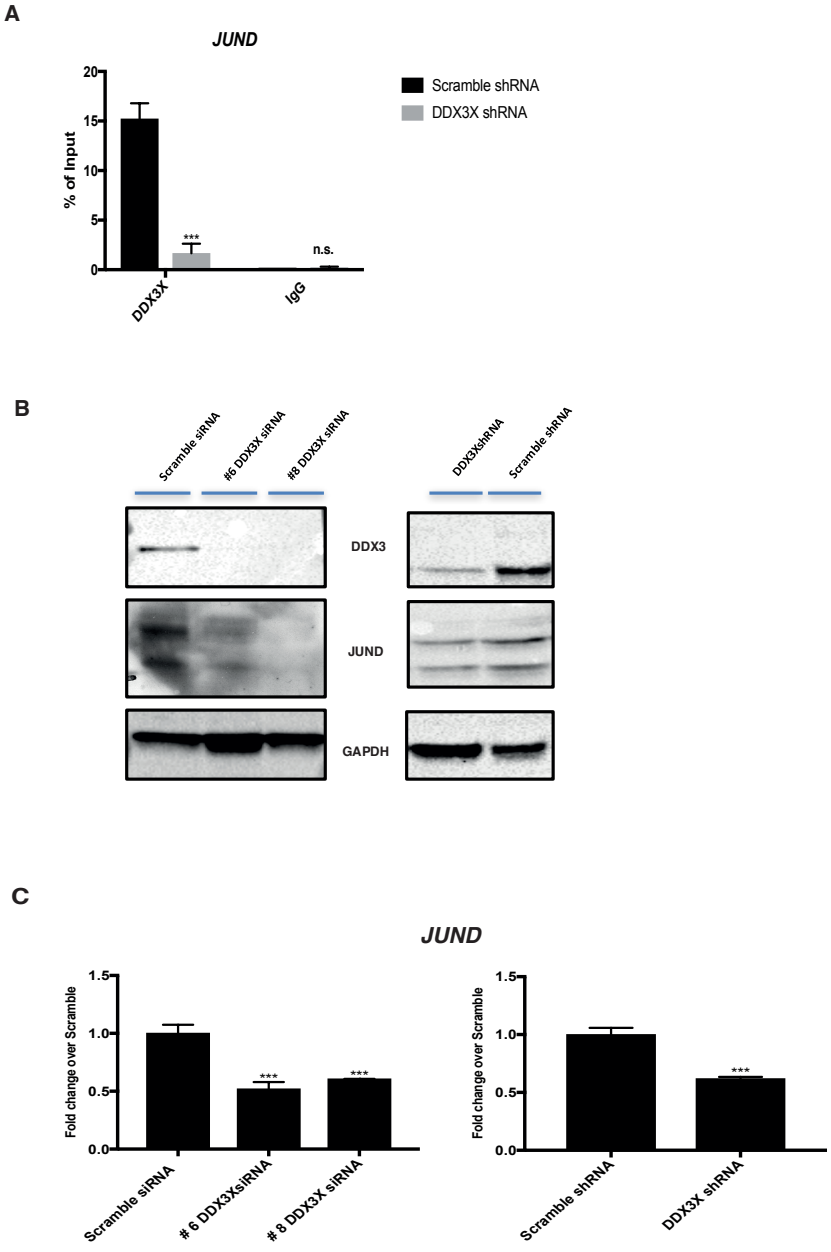


Figure 4.13. JUND expression is affected by DDX3X knock down.
(A) DDX3X and IgG CLIP-qPCR were performed to test the binding of *JUND* mRNA in MCF7 cells transduced with either scrambled shRNA or DDX3X shRNA. cDNA obtained from each immunoprecipitation was analysed by qPCR to determine enrichment of the indicated targets. (B) Western blot shows DDX3X, JUND and GAPDH protein levels in whole cell lysates of MCF7 cells transduced with either scrambled shRNA or two different DDX3X shRNA (right panels) and in MCF7 cells transfected with either scrambled siRNA or two different DDX3X targeting siRNAs (#6 or #8) (left panels). (C). mRNA levels of *JUND* under the same conditions as in B measured by qPCR. (A,C) Results represent the average of three biological replicates (Section 2.5.3). Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble controls (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001; **** ≤ 0.0001).

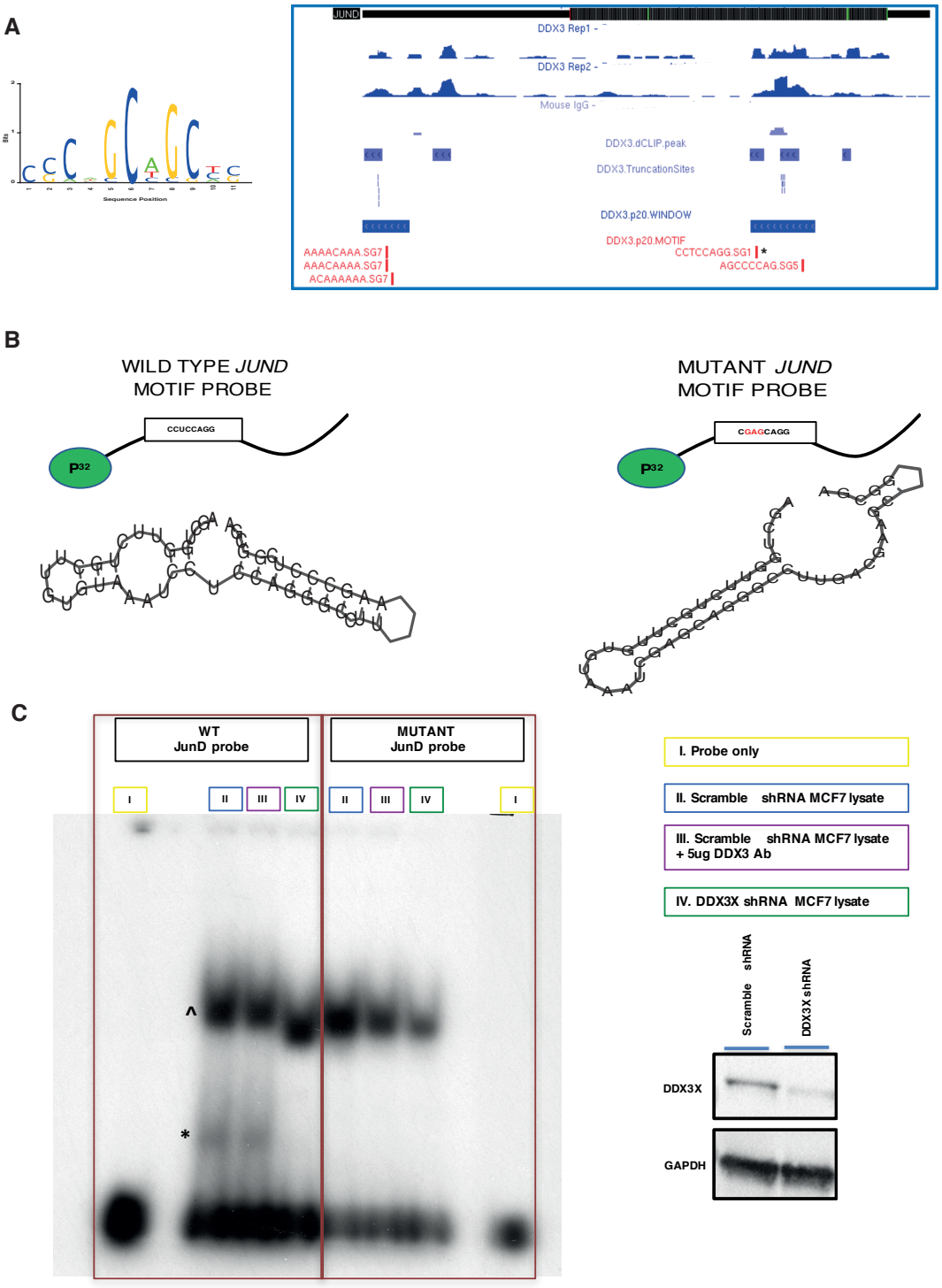


Figure 4.14. DDX3X binds the human JUND transcript on a discrete sequence motif.
(A) DDX3X consensus sequence of the binding motif found in *JUND* mRNA (left panel). Visualization of the USCS custom tracks for DDX3X iCLIP truncation

sites and motif localization on *JUND* sequence (left panel-validated motif is marked with *). (B) Model of the wild type *JUND* and mutant *JUND* RNA motif-containing probes used in EMSA assays; Minimum Free Energy (MFE) secondary structure of the RNA oligonucleotide probes used in EMSA, predicted using RNAfold tool (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi) (C) Shown is the autoradiograph of the mobility shift assay performed as described (Section 2.7). Briefly, whole cell lysate from MCF7 cells transduced with either scrambled shRNA (lanes II and III) or *DDX3X* targeting shRNA (lanes IV) were used to test the binding of either wild type (left half) or mutant (right half) *JUND* probe. Anti-DDX3X antibody was added (lanes III). Probe signal (lanes I). (D) Western blot shows protein level of DDX3X and GAPDH in whole lysate of MCF7 cells used in C.

4.8 DDX3X binds target RNAs that share similar predicted structures.

It is possible that structural features of target RNAs mediate interactions with binding proteins. I therefore decided to investigate whether DDX3X target RNAs, so far identified and validated, share common predicted structures at the region bound by DDX3X. To do this, I considered an extended region of approximately 200 nucleotides surrounding the first iCLIP site located at the 5′-side of the *RPS2, RPS3, RPS6* and *JUND* transcripts (marked with * in Figure 4.12).

I used the locARNA tool to compare the 2D structure of selected RNAs. The software first generates a phylogenetic tree based on the alignment of the sequences. As shown in Figure 4.15A, *RPS6* sequence has lower homology when compared to the other RNAs sequences, whereas *RPS2, RPS3* and *JUND* show higher homology among them. I therefore decided to run comparative structural prediction on *RPS2, RPS3* and *JUND* sequences, to improve the quality of the prediction results. LocARNA generates a structural prediction combining all potential common structures of the aligned RNA sequences. The outputs of locARNA are shown in Figure 4.15B (sequence and structure alignment) and C (overlaid structure of three RNAs). The sequence and structure of the three RNAs are well aligned showing strong conservation in forming stem-loop structures (Figure 4.15B). More interestingly, DDX3X iCLIP binding sites are located in three consecutive stem-loop structures (highlighted in Figure 4.15C). In the case of *RPS3* mRNA, the region used in the alignment contains two consecutive distal binding sites that both lie on different stem loop structures (Figure 4.15C). Importantly, one of the *RPS3* binding sites is located within the same stem-loop as the two consecutive *JUND* binding sites (Figure 4.15C). These results suggest that DDX3X binding to its target RNAs may preferentially occur at stem loop RNA secondary structures.

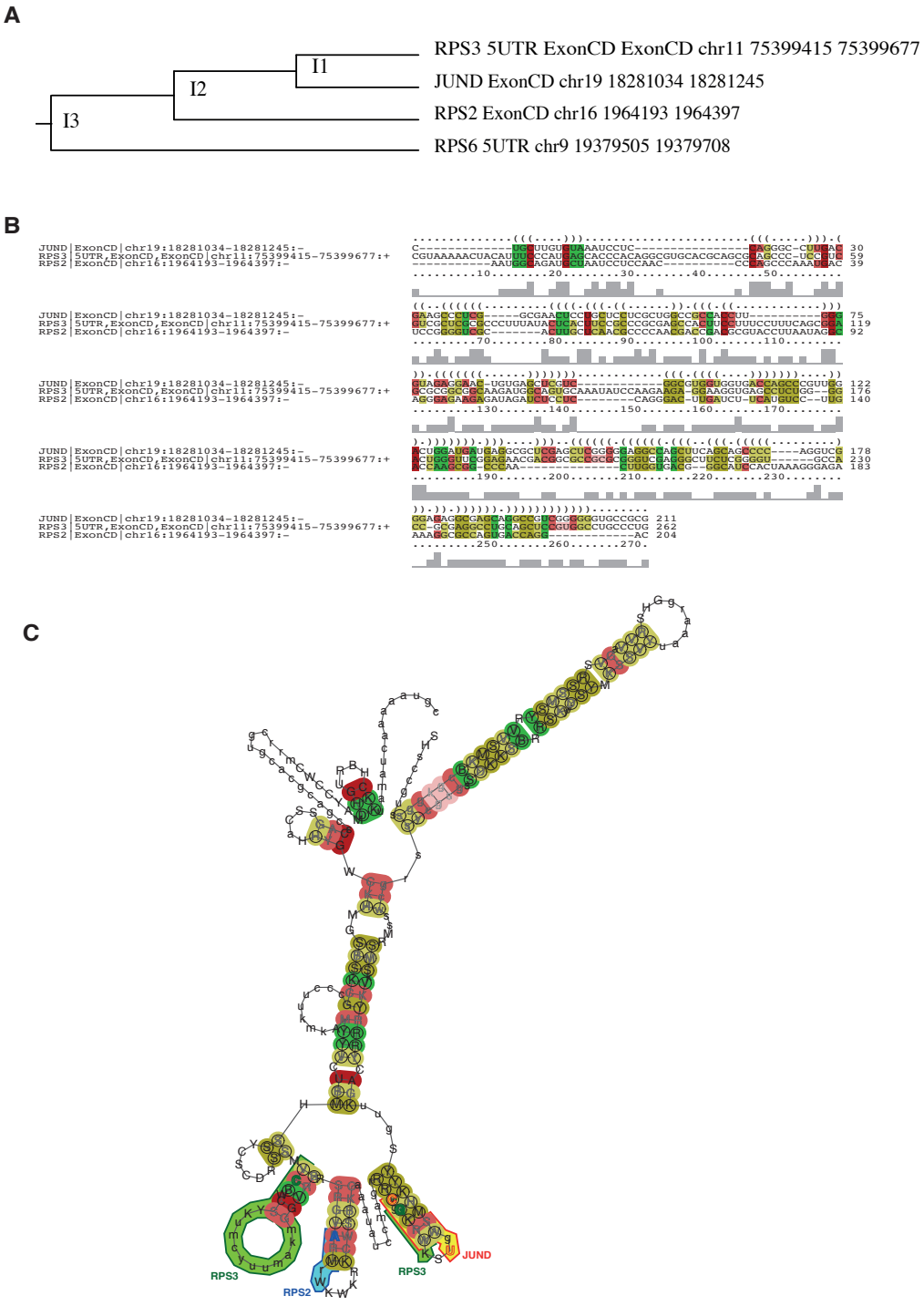


Figure 4.15. DDX3X binds its target RNAs with similar secondary structures.
(A) Phylogenetic tree generated with locRNA [PMCID: PMC3334699] by alignment of *RPS2*, *RPS3* and *JUND* sequences (indicated by the genomic coordinates) extended around iCLIP binding sites.
(B-C) locRNA output of sequence and structure alignment (B) and overlaid structure prediction (C) of the three RNA sequences. (B-C) Colour code indicates the number of possible base pairs (red=1; yellow=2; green=3); (C) Sequence in the overlaid structure is shown in IUPAC.

4.9 Discussion

Implementation of iCLIP to discover direct target RNAs of DDX3X uncovered a wealth of information that was useful for downstream investigation into the biological roles and mechanistic features of this helicase in MCF7 breast cancer cells. The results of this chapter provide insight into the roles of DDX3X and indicate possible important functions for this protein. These possibly include maintaining RNA stability and promoting translation of target RNAs. Moreover, I was able to show preferential binding of DDX3X to a specific sequence motif within the *JUND* human transcript.

4.9.1 DDX3X regulates expression of its target RNAs and recognizes at least one relevant sequence motif.

DDX3X regulates several aspects of RNA metabolism (Section 1.5). Recently, DDX3X CLIP-seq data sets, generated from HEK293T cells over-expressing either wild-type or medulloblastoma-associated catalytic mutants of DDX3X, were published [203]. Interestingly, these data show enrichment of DDX3X binding sites in coding sequences of mainly translation factor mRNAs, which is in strong agreement with my findings. Also, over-expression of DDX3X mutants led to aberrant formation of stress granules and caused global translational impairment [203].

I found that depletion of DDX3X affects the levels of many of its direct target RNAs as well as the proteins they encode. This suggests that the role of DDX3X may not be solely involved with translational regulation at the ribosome level. My findings suggest that the global translational impairment of DDX3X mutant expressing cells [203], as observed by Valentin-Vega and colleagues, may be caused by lack of DDX3X positive regulation of mRNAs encoding translation-related proteins (such as ribosomal proteins), which ultimately results in generally defective translation.

Binding of DDX3X to the 5′-cap of mRNAs together with other translational initiation factors is dependent upon mTOR activity [96]. Proteomic studies using m^7GTP affinity chromatography showed that presence of DDX3X in the 5′-cap-binding complex is dependent on the activation state of mTOR. Specifically, the amount of DDX3X interacting with m^7GTP increased when mTORC1 signalling was activated by insulin treatment, and it was reduced when mTORC1 signalling was inhibited by PI3K inhibitors [96]. Based on these published findings, as well as on the pathway enrichment in my DDX3X iCLIP gene ontology (Figure 4.7A) which highlighted the mTOR pathway term, I used rapamycin to test the hypothesis that the activity of DDX3X might be linked to the mTOR signalling pathway. Rapamycin inhibits the kinase activity of mTOR thereby blocking the signalling cascade that leads to protein synthesis and other processes essential for cell growth (Figure 4.7B-C). As such, it is widely used to study events downstream of mTOR [105] [88, 107].

DDX3X binding to target mRNAs encoding *RPS2, RPS3* and *RPS6* was not affected by short time-course inhibition of mTOR by rapamycin (Figure 4.8), and neither were the expression levels of the ribosomal mRNAs and proteins. Conversely, DDX3X's binding to the same mRNAs remained fairly constant after 24h of rapamycin treatment (Figure 4.9). At the same time, DDX3X target RNAs levels were slightly elevated. This enhanced expression appears to be independent of DDX3X since a similar increase is seen in DDX3X knock down cells after 24h of rapamycin treatment. On the other hand, the protein levels of RPS2, RPS3 and RPS6 were differentially affected by rapamycin. In particular, reduction of DDX3X, RPS2 and RPS3 protein levels upon rapamycin treatment likely occurs via a DDX3X-independent mechanism.

Taking all of the previous observations into consideration, my results indicate that, in conditions that phenocopy metabolic stress, such as the inhibition of the mTOR pathway, which also reduce DDX3X levels, the helicase binds preferentially to its target RNAs presumably at the expenses of other DDX3X functions. The functional role of this binding

does not seem to be general translational regulation. In fact, if DDX3X is directly part of the translational apparatus that positively responds to the mTOR pathway, binding of DDX3X to its target RNAs would likely decrease in response to rapamycin.

Ribosomal proteins and other translation factor mRNAs share a specific sequence motif called TOP (terminal oligopyrimidines or track of pyrimidines) sequence localized at their 5′-UTR. In these mRNAs, the 7-methyl G cap is followed by a cytosine residue and a oligopyrimidine tract of 7-14 residues that forms the 5′-TOP motif. It is known that 5′-TOP-containing RNAs are translationally regulated by mTOR since they have been shown to be among the most sensitive targets of mTOR inhibition. The mechanism through which 5′-TOP mRNAs are translationally co-regulated by the mTOR pathway remains poorly understood (Section 1.1.3). I found that most of the ribosomal protein RNAs in the DDX3X CLIP dataset bear a pyrimidine rich motif at their DDX3X binding sites (Figure 4.11C-D). However, I did not observe enrichment of DDX3X binding at the 5′-UTRs of target RNAs. Furthermore, the ribosomal protein target RNAs showed the presence of different DDX3X binding motifs in many different regions of the transcript (Figure 4.12). This indicates that DDX3X binding to ribosomal proteins mRNAs and other targets may involve a "TOP like" motif in terms of sequence composition, but not necessarily in terms of functionality. Infact, this is consistent with the lack of evidence for a direct role of DDX3X RNA binding on regulation of proteins levels in the mTOR pathway (Section 4.5).

Interestingly, one of the motifs found in the *JUND* transcripts coding sequence was also a CT (i.e. CU in RNA) rich motif (Figure 4.14A). The results from the EMSA assay suggest the presence of a protein complex in MCF7 cells that contains and requires DDX3X to bind *JUND* mRNA at that specific sequence. In fact, a band with high mobility is only observed when MCF7 cell lysates expressing DDX3X are used together with a *JUND* wild type motif probe.

A band with lower mobility very likely represents binding of a protein complex to a region of the *JUND* probe distinct from the motif site. However, DDX3X may well be a component of this complex as the band's mobility is increased when extracts from DDX3X depleted cells are used with the wild type *JUND* probe. A band of similarly lower mobility is still observed when MCF7 cell lysates are used together with a *JUND* mutant motif probe. The presence of DDX3X in this complex has presumably a co-operative effect for the binding of other putative factors to the regions outside the relevant motif, as the depletion of DDX3X from MCF7 lysates results in reduced intensity of the low mobility band when using the *JUND* mutant probe. Unfortunately, I could not validate the presence of DDX3X in the complex that binds the *JUND* probe in my supershift EMSA experiments. This could be due to several technical issues such as the inefficiency of the antibody to perform supershift or being stable at the gel electrophoresis conditions. Also, the DDX3X epitope recognized by the antibody could be masked by the protein-probe interaction. In the future, the possibility of attempting supershift with an anti-DDX3X antibody that recognizes a different epitope, could be explored.

Notably, *RPS2* mRNA bears a DDX3X binding motif identical to the one found and validated in the *JUND* sequence (motif SG1 CCTCCGAG – Figure 4.14A; Figure 4.12). In contrast, *RPS3* and *RPS6* mRNAs bear multiple DDX3X binding sites (mostly CA or GT rich, respectively), with low similarity to the ones in *JUND* and *RPS2* sequences (Figure 4.12). Like JUND, RPS2 protein was strongly depleted upon DDX3X knock down and this effect was more prominent than observed for RPS6 (Figure 4.6A – Figure 4.13B). Interestingly, RPS3 mRNA was reduced upon DDX3X depletion but the level of RPS3 protein was unaffected.

These exciting observations suggest that the binding of DDX3X to different motifs may affect the translational fate of the relevant mRNAs. Although translational studies (i.e. polysome profiling) are necessary to define whether DDX3X is effectively regulating the translation of its

bound targets through this motif, one can speculate that DDX3X binding to specific motifs may dictate recruitment of different factors to its targets or, conversely, DDX3X may be recruited to specific sequences by the interaction with proteins exerting different functions (such as translation rather than degradation).

My findings indicate that DDX3X preferentially binds stem-loop RNA structures. Indeed, in the comparative structural prediction analysis I found that DDX3X binding sites on *RPS2, RPS3* and *JUND* mRNAs lie on three consecutive stem loops. Importantly, the binding site is predicted to localize at the edge between the single strand and the double strand portions of the stem loop, which is entirely consistent with the mechanism of loading of DEAD box RNA helicases onto their targets: as described in the introduction chapter (Section 1.3) these proteins use a single strand extension to be loaded on to the target double strands they subsequently unwind.

By combining all the evidence regarding the possible functional role of the binding motif and the latter observations on the structural binding preferences of DDX3X, the overall results of this chapter suggest that the binding of DDX3X may be directed by RNA substructures and could be relevant for the stability of the bound mRNAs. However, the functional consequences of binding may depend on the specific sequence bound, which could call into action other proteins to exert differential functions on the RNA.

Chapter 5 - DDX54 and DDX3X as cell cycle regulators in breast cancer MCF7 cells

5.1 Introduction

DDX RNA helicases play essential roles in the biology of cells of very different tissues and they have been associated with the onset of several different cancers (Section 1.3). As described in Section 1.6, the roles of DDX54 and DDX3X in breast cancer *in vitro* models have been previously reported, although detailed investigations of these helicases biological function in breast cancer cells was not shown.

Having studied mechanistic features of DDX3X and DDX54 function via their iCLIP RNA binding profile, I next investigated the biological roles of these proteins in MCF7 breast cancer cells, as well as in immortalized non-tumorigenic MCF10A breast epithelial cells.

5.2 Aims of the chapter

This chapter had the following aims:

- To determine the effects of knocking down DDX54 and DDX3X on cellular proliferation in MCF7 cells.
- To determine gene expression changes upon DDX54 or DDX3X knock down
- To analyse cell cycle changes upon DDX54 or DDX3X knock down
- To provide mechanistic insight into the roles of DDX54 and DDX3X through the intersection of iCLIP data with gene expression changes upon helicases knock down.

5.3 Effects of DDX54 and DDX3X knock down on proliferation of immortalized breast epithelial cells and transformed breast cancer cells

To determine whether the roles of DDX54 and DDX3X are essential in transformed breast cancer cells, but not in non-tumorigenic immortalized breast cells, I performed proliferation assays in MCF7 cells (transformed) and MCF10A cells (immortalized) upon knock down of DDX54 or DDX3X helicases. Cells were transfected with either scramble siRNA or two different siRNAs targeting either DDX54 or DDX3X. 24 h after transfection (day 0) cells were seeded for proliferation assays and counted at the indicated time-points up to 6 days after seeding (Figure 5.1A-B).

Figure 5.1 shows the sensitivity of MCF7 and MCF10A cells to knock down of DDX54 (panel A) and DDX3X (panel C). Panels A and C show the number of living cells counted at each indicated time-point, panels B and D show depletion of DDX54 and DDX3X upon siRNA-mediated knock down, at the indicated time-points. Proliferation of both MCF7 and MCF10A cells showed sensitivity to DDX54 knock down (Figure 5.1A). As shown by the western blot in Figure 5.1B, the knock down efficiency of DDX54 protein was different in the two cell lines. In MCF10A cells, virtually no recovery of DDX54 expression occurred up to day 6. At 72h after transfection (day 2), when the levels of DDX54 depletion were comparable between the two cell lines, the sensitivities of MCF7 and MCF10A cells proliferation to DDX54 knock down were similar, with both cell lines showing basically no cell growth. In contrast, scramble siRNA-transfected control cells grew normally.

MCF7 and MCF10A cells show differential sensitivity to DDX3X knock down (Figure 5.1C). As can be seen in Figure 5.1D, recovery of DDX3X expression during the time course was faster in MCF7 cells. At 72h after transfection (day 2), when depletion of DDX3X was virtually complete in both cell lines, the depleted MCF7 cells showed no growth compared to

scramble control cells. In contrast, DDX3X depleted MCF10A cells displayed a growth rate similar to scramble control cells. Indeed, MCF10A cells showed almost no difference in proliferation to the scramble control cell line at any time-point, irrespective of the expression level of DDX3X (Figure 5.1C-D).

Overall, these findings indicate that DDX54 is essential for proliferation of transformed breast epithelial cancer (MCF7) and immortalized non-tumorigenic breast epithelial (MCF10A) cells. Conversely, DDX3X is mainly required for the growth of transformed MCF7 cells.

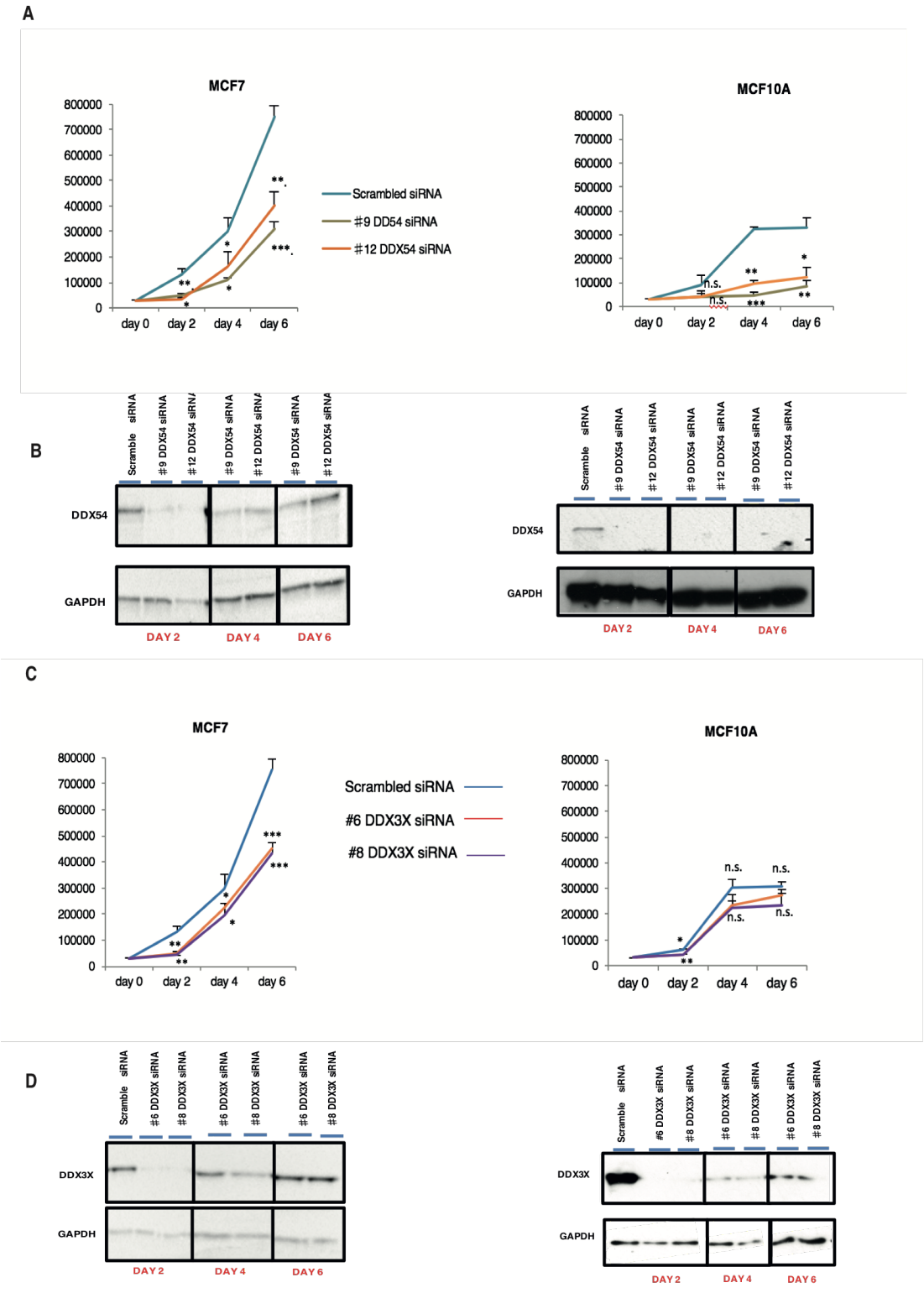


Figure 5.1. DDX3X is specifically required for breast cancer cells proliferation. (A and C) Cell viability curves for MCF7 (left panels) and MCF10A (right panels) cells. Cells were transfected with either scrambled siRNA or two different siRNAs targeting either DDX54 (#9 or #12) (A) or DDX3X (#6 or #8) (C) 24h after transfection, cells were seeded for proliferation assays and counted at the

indicated time-points up to 6 days after seeding. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001). (B and D) Western blots showing the protein levels of DDX54 (B), DDX3X (D) and GAPDH (B and D) in MCF7 (left panels) and MCF10A (right panels) transfected with either scrambled siRNA or the two different siRNAs targeting each helicase. Protein levels were measured at each indicated time-point corresponding to counting time-points.

5.4 DDX54 and DDX3X regulate MCF7 cell cycle

To gain further insight into the biological roles of DDX54 and DDX3X in MCF7 cells, and the pathways linked to reduced proliferation observed upon depletion of these proteins, I performed RNA-seq analysis of MCF7 cells following knock down of DDX54 and DDX3X. MCF7 cells were transfected with either scrambled siRNA or two different siRNAs targeting DDX54 or DDX3X. 72h after transfection, RNA was isolated from cells and used for Illumina RNA-seq library preparation (Section 2.11). Knock down of both proteins caused both down-regulation and up-regulation of genes (Figure 5.2) with good reproducibility between the two siRNAs used for each protein (Figure 5.3A- Appendix Figure 3).

Strikingly, when I cross-compared gene expression changes caused by the depletion of DDX54 versus DDX3X, I found overlap of downregulated genes between the two datasets (Figure 5.3B). This suggests the two proteins might have partially redundant function in gene regulation.

I then performed a gene ontology search using DAVID functional annotation tool to categorize the differentially expressed genes in both DDX54 and DDX3X knock down cells. Figure 5.4 shows the GO terms related to biological processes (panel A) and cellular components (panel B) for genes down-regulated upon DDX3X (left) and DDX54 (right) knock down in MCF7 cells. This gene ontology analysis highlights a striking resemblance in biological processes and cellular component categories. In particular, it indicates that genes encoding proteins involved in cell cycle and nuclear components were down-regulated upon knock down of either DDX3X or DDX54 in MCF7 cells.

Following on from the above observation, I next analysed cell cycle progression of MCF7 cells when either DDX54 or DDX3X was knocked down (Figure 5.5, Figure 5.6). Cells were transfected with scramble siRNA or with two different siRNAs against either DDX54 or DDX3X. 24h after transfection, cells were seeded for cell cycle analysis. 48h after transfection, cells were treated overnight with 50 ng/ml nocodazole and tested by propidium iodide (PI) staining in flow cytometric analysis the

next day. Nocodazole is a chemical compound that interferes with the polymerization of microtubules and therefore with the formation of the mitotic spindle. Nocodazole treatment blocks cells at the G2/M transition allowing a comparative analysis of cell cycle from cell populations in different conditions equally forced to arrest in G2/M. Figure 5.5A and Figure 5.6A show the proportion of cells in each cell cycle phase (G1, S, G2/M) normalized to scramble control cell values. Figure 5.5B and Figure 5.6B show depletion of either DDX54 or DDX3X proteins in the tested cells at the time point when the cells were harvested for cell cycle analysis. Figure 5.5C and Figure 5.6C show representative flow cytometric images of PI staining analysis of the respective graphs (Figure 5.5A and Figure 5.6A). As shown, both knock down of DDX54 (Figure 5.5) and DDX3X (Figure 5.6) caused a dramatic cell cycle arrest in G1.

These data, together with the changes in gene expression quantified by RNA-seq, indicate that both DDX54 and DDX3X exert regulatory activity on the cell cycle of MCF7 cells, specifically impairing progression of cells through S phase.

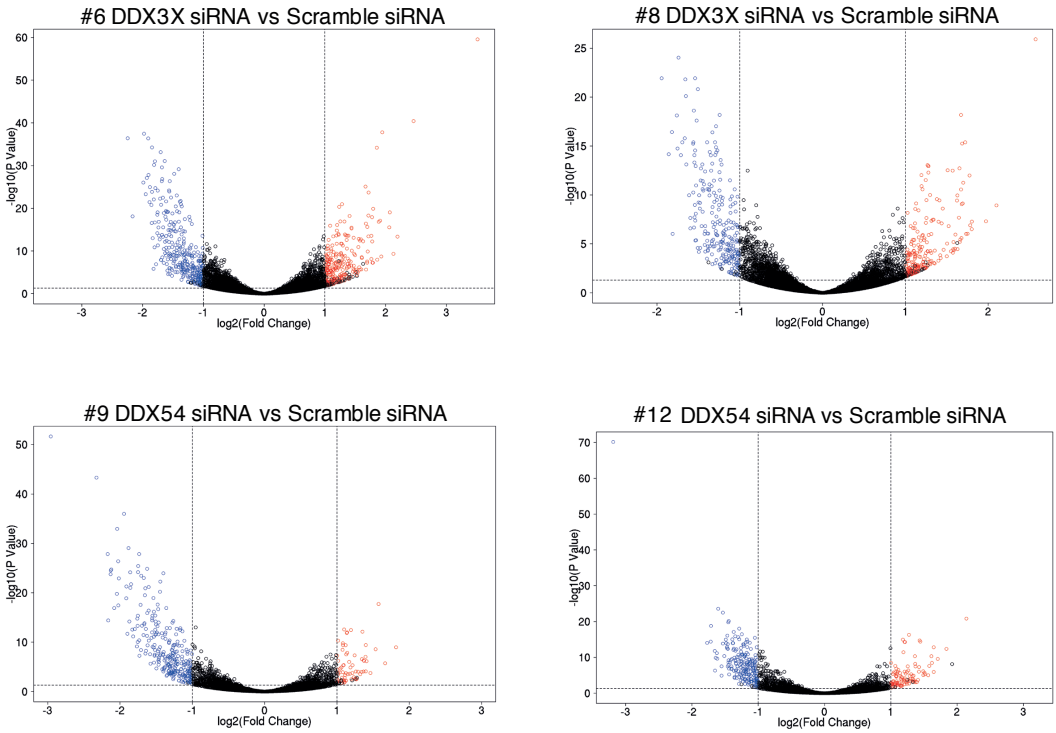


Figure 5.2. Gene expression changes upon knock down of DDX3X and DDX54. Volcano plots for each individual differential expression analysis. Significant genes (based on the specified threshold values) are indicated by the red (up-regulated) and blue (down-regulated) points of the plot. Genes that are not changing in either condition are coloured black.

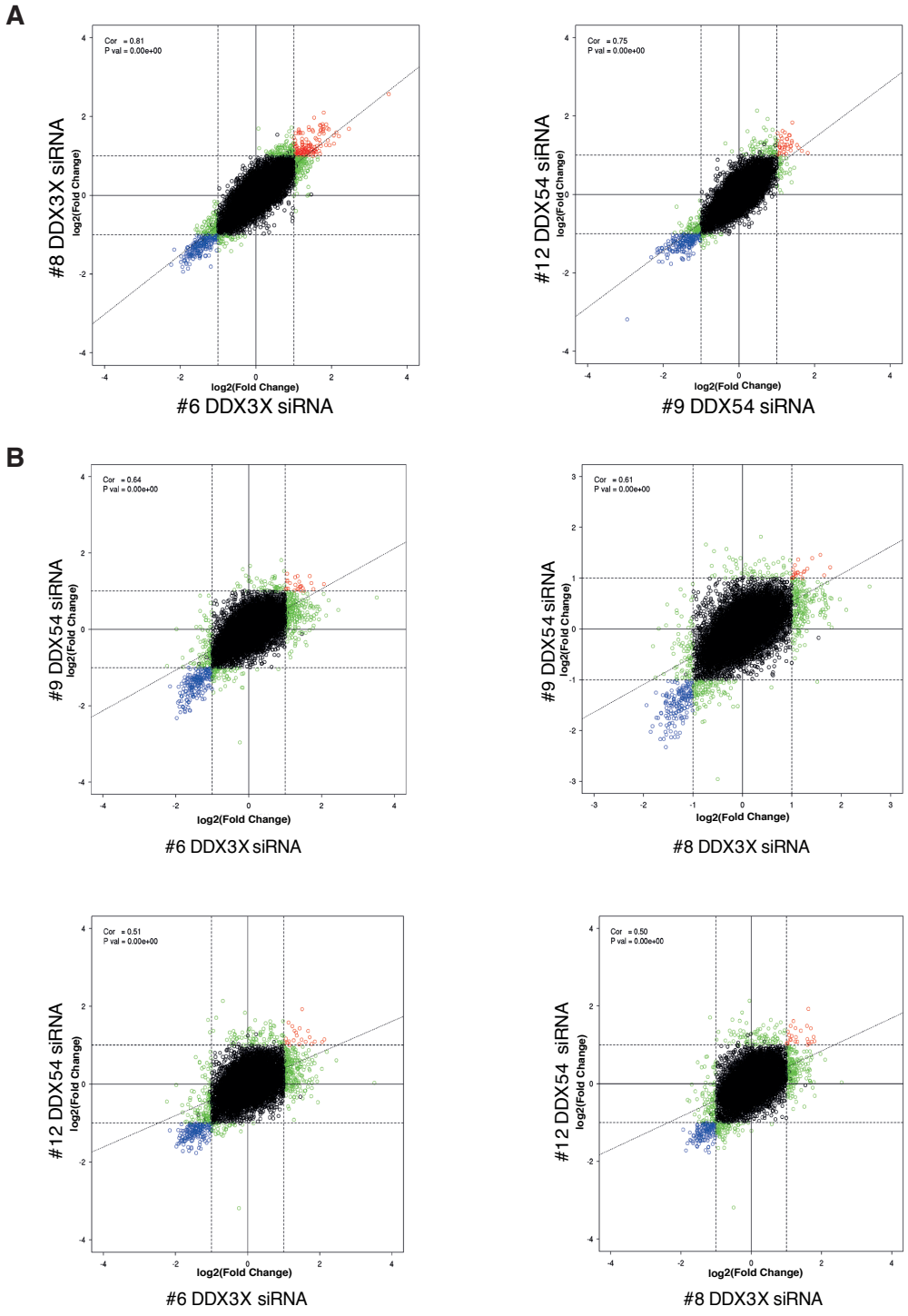


Figure 5.3. Correlation of gene expression changes amongst knock down samples. Correlation plots between the log-fold change for the various comparisons. Scatter plots show the log-fold change for one siRNA versus another targeting the same protein (A) or either DDX3X or DDX54 for cross-comparison (B). Colour code indicates whether or not they are up-regulated in both conditions (red), down-regulated in both conditions (blue), significant in one condition or the other (green), or they are not significant (black). The corresponding Pearson correlation ("Cor") and p-value for a linear model ("P val" and thin dotted line) are shown.

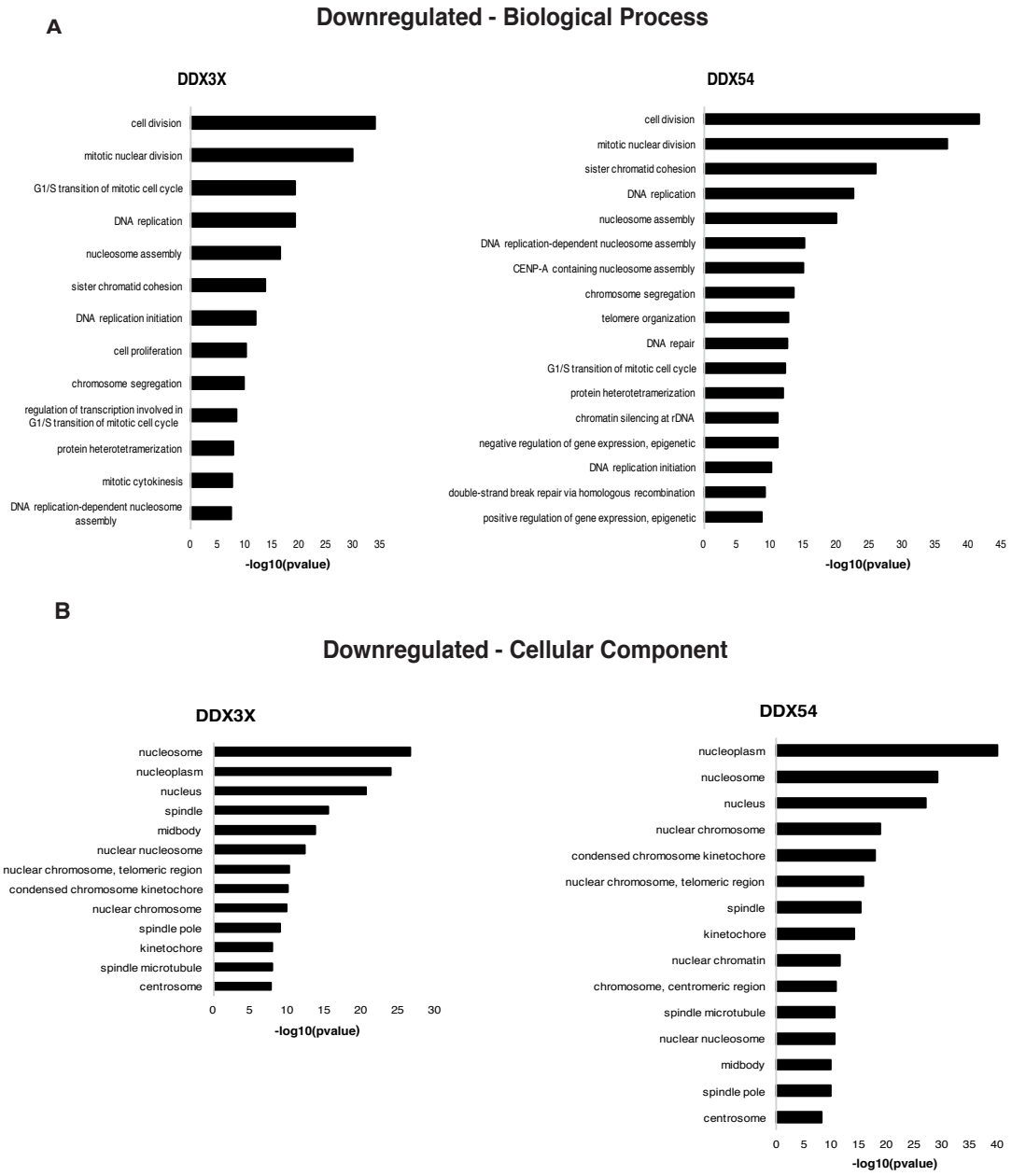


Figure 5.4. GO term analysis of down-regulated genes upon knock down: DDX3X and DDX54 are cell cycle regulators.

Down-regulated genes in knock down DDX3X or DDX54 MCF7 cells were identified as described in Methods (Section 2.11.2) from RNA-seq data (Section 2.11.1). The GO enrichment of down-regulated genes was obtained using the DAVID functional annotation tool with default settings [PMID: 19131956]. The p-values were calculated using the "default" method of DAVID. (A) Y-axis shows the GO terms related to biological processes in DDX3X knock down MCF7 cells (left panel) and in DDX54 knock down MCF7 cells (right panel); (B) Y-axis shows the GO terms related to cellular components in DDX3X knock down MCF7 cells (left panel) and in DDX54 knock down MCF7 cells (right panel); (A-B) X-axis shows –log10 p-values of each GO terms.

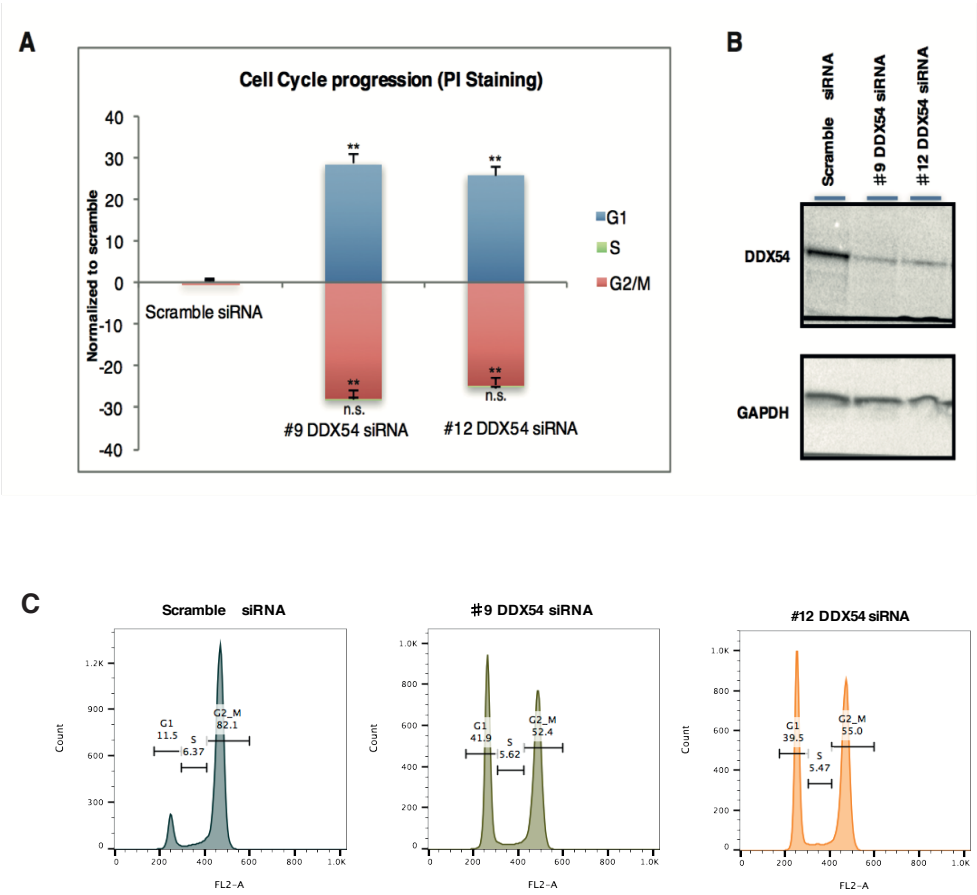


Figure 5.5. DDX54 knock down arrests MCF7 cell cycle in G1.
MCF7 cell cell cycle progression flow cytometric analysis based on PI staining of MCF7 cells 72h after transfection with either scrambled siRNA or two different DDX54 (#9 or #12) targeting siRNAs. (A) Fold change quantification for each phase (G1-S-G2M). Results represent the average of two biological replicates. (B) Western blots show the protein levels of DDX54 and GAPDH in MCF7 cells transfected with scrambled siRNA or with DDX54 (#9 or #12) targeting siRNAs. (C) Images from FlowJo analysis of flow cytometric analysis of cell cycle progression upon PI staining as quantified in A (images represent one replicate). Cells were treated with 50 ng/ml nocodazole overnight. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001; **** ≤ 0.0001).

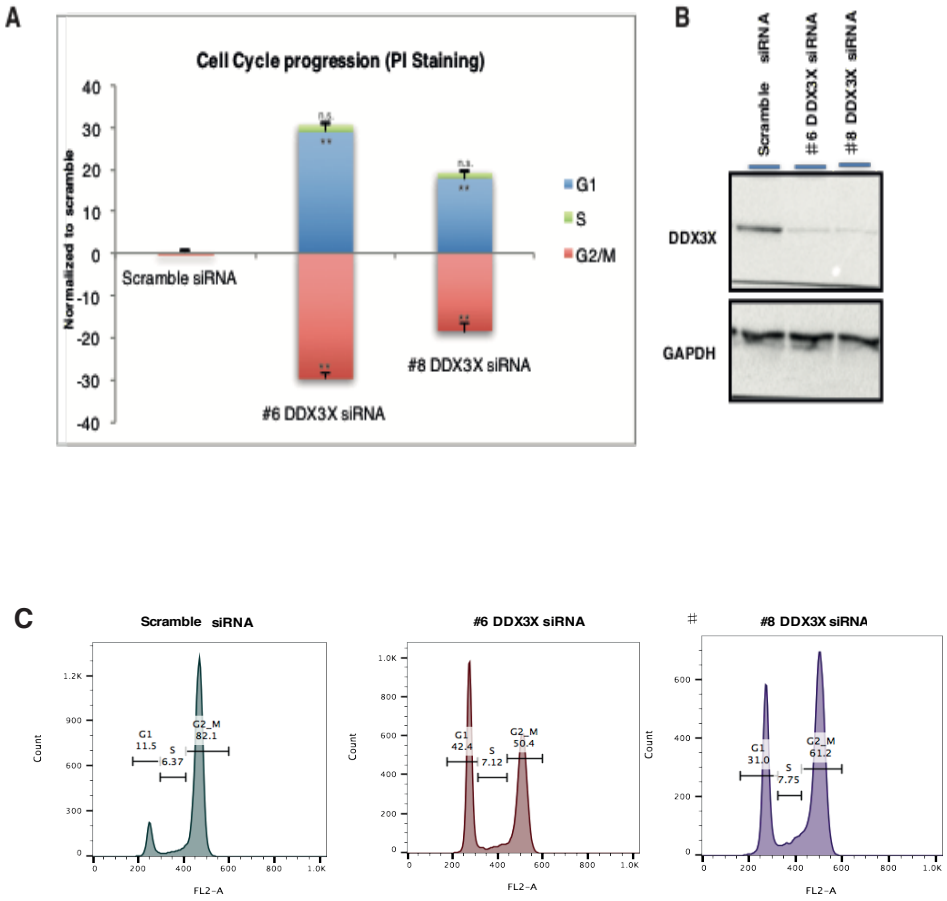


Figure 5.6. DDX3X knock down arrests MCF7 cell cycle in G1.
MCF7 cell cell cycle progression flow cytometric analysis based on PI staining in MCF7 cells 72h after transfection with either scrambled siRNA or two different DDX3X (#6 or #8) targeting siRNAs. (A) Fold change quantification for each phase (G1-S-G2M). Results represent the average of two biological replicates. (B) Western blots show the protein levels of DDX3X and GAPDH in MCF7 cells transfected with scrambled siRNA or with DDX3X (#6 or #8) targeting siRNAs. (C) Images from FlowJo analysis of flow cytometric analysis of cell cycle progression upon PI staining as quantified in A (images represent one replicate). Cells were treated with 50 ng/ml nocodazole overnight. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

5.5 DDX3X affects cell cycle through regulation of KLF4 transcription factor

Transcriptome and cell cycle analysis in cells depleted of DDX54 and DDX3X highlighted the role of both proteins in cellular proliferation and cell cycle progression in breast cancer cells. This function might be redundant in these cells, as knock down of either of the two proteins results in down regulation of a similar set of genes (Figure 5.3, Figure 5.4). However, the activity of DDX3X seems to be specific for transformed breast cancer cells, as knock down for DDX3X has a negligible effect on proliferation of immortalised non-tumourigenic MCF10A breast epithelial cells (Figure 5.1). In contrast to DDX54, this suggests a possible selectivity of DDX3X towards a transformed cancer phenotype. For this reason, I decided to focus on DDX3X for the following validation of RNA-seq data.

To this purpose, I performed qPCR validation of a number of cell cycle relevant target genes whose expression was altered in DDX3X knock down MCF7 cells (Figure 5.7). The table in Figure 5.7A includes the genes tested and indicates how their expression was changed in the RNA-seq analysis upon DDX3X knock down. Figure 5.7B shows the mRNA levels of *CCNA2*, *CCNE1, CDK2* and *KLF4* genes in normal and DDX3X knock down MCF7 cells. Reassuringly, the expression of these genes was altered as predicted from the RNA-seq data analysis: upon DDX3X knock down, *CCNA2* and *CDK2* expression was reduced, *CCNE1* expression was unchanged and *KLF4* expression was increased. As further validation, I also showed that the protein level of CDK2 was reduced in MCF7 cells upon DDX3X knock down (Figure 5.7C).

Next, I tested whether the cell cycle defects observed in MCF7 cells upon DDX3X knock down could be linked to DDX3X RNA binding activity. To this end, I intersected the DDX3X knock down RNA-seq dataset with the DDX3X iCLIP dataset to search for cell cycle-related common targets that would suggest binding of DDX3X is necessary for their expression. As shown in Figure 5.8, the overlap between these two datasets was very

limited, indicating that the RNA level of most of the direct DDX3X target RNAs was not significantly affected by DDX3X knock down. Interestingly however, one of the targets common to the two datasets was KLF4. *KLF4* RNA was bound by DDX3X, according to iCLIP, and it was up-regulated in DDX3X knock down MCF7 cells, as determined by RNA-seq and validated by qPCR (Figure 5.7B).

To further explore the interaction between DDX3X and KLF4 RNA, I validated binding of DDX3X to *KLF4* mRNA by performing CLIP-qPCR, as previously described (Section 2.13.3), with anti-DDX3X antibodies as well as anti-DDX24 and anti-IgG antibodies as controls. *KLF4* mRNA was only precipitated using the anti-DDX3X antibody and was absent in anti-DDX3X immunoprecipitation performed in DDX3X knock down MCF7 cells (Figure 5.9A) indicating specificity of DDX3X binding. To investigate the function of DDX3X binding on KLF4 mRNA, I tested the KLF4 protein levels in DDX3X knock down MCF7 cells and observed up-regulation of the protein. This correlates well with up-regulation of KLF4 mRNA upon DDX3X knock down (Figure 5.9B).

KLF4 is a transcription factor and a negative regulator of the G1/S transition in the cell cycle[248]. Therefore, I tested the hypothesis that DDX3X negatively regulates the protein level of *KLF4,* which in turn would promote cell cycle progression. I decided to focus on the link between KLF4 up-regulation and down-regulation of cyclins and cyclin-dependent kinases observed upon DDX3X knock down (Figure 5.7B), whose repression would ultimately result in the observed G1 arrest.

If KLF4 and DDX3X actions are linked in determining the phenotype observed, then depletion of KLF4 should increase the expression of cyclin and cyclin-dependent kinase genes down-regulated upon knock down of DDX3X. Conversely, *KLF4* over-expression should phenocopy the effects of knocking down DDX3X. Therefore, I performed both KLF4 knock down

and over-expression experiments in MCF7 cells. Cells were transfected with either scramble siRNA or KLF4 specific siRNA. 72h after transfection, cells were collected and RNA was isolated in order to analyze expression of specific genes. As shown in Figure 5.9C, knock down of KLF4 enhanced expression of *CCNA2* and *CDK2*, whilst *CCNE1* and *DDX3X* gene expression remained unchanged.

Figure 5.9D and E show how KLF4 over-expression in MCF7 cells affects expression of the same genes and MCF7 cell cycle, respectively. MCF7 cells were transfected with increasing amounts of a plasmid over-expressing the human isoform of *KLF4* (pPB-CAG-hKLF4-pA-pgk-Hygro). 24h after transfection, cells were seeded for cell cycle analysis. Cells were treated overnight with 50 ng/ml nocodazole and tested for PI staining in flow cytometric analysis the next day. At the same time-point, cells were harvested for gene expression analysis by qPCR. Over-expression of *KLF4* caused an increase in the fraction of G1 arrested cells proportional to the amount of h*KLF4* transfected (Figure 5.9E). Consistent with this, expression of the genes that were up-regulated upon DDX3X knock down (*CCNA2, CDK2*) was increased proportionally to the extent of *hKLF4* over-expression. Conversely, *CCNE1* gene expression, which was unchanged in DDX3X knock down, did not show any change upon KLF4 over-expression (Figure 5.9D).

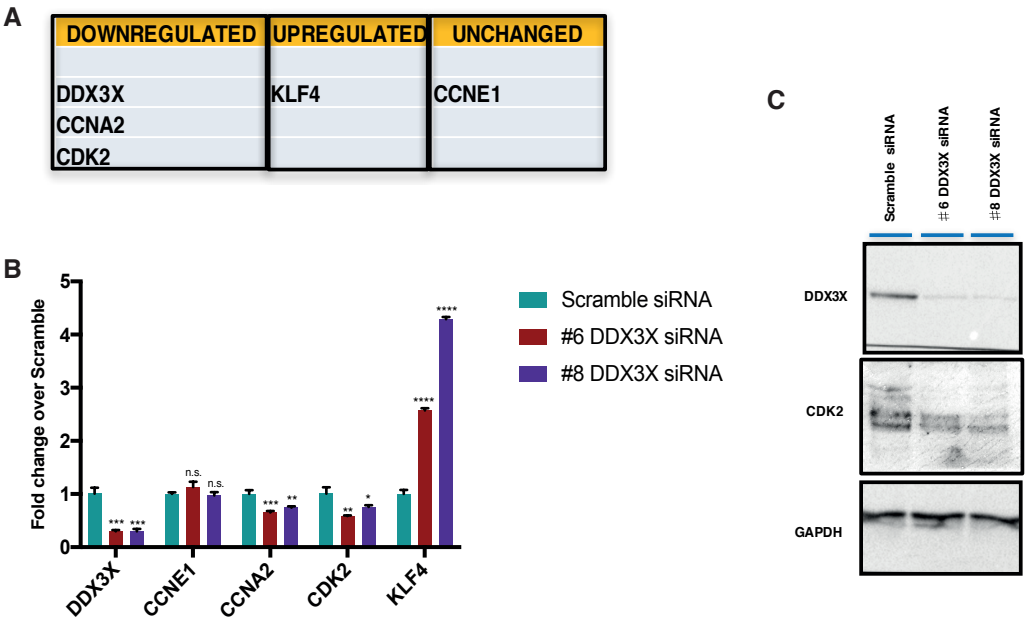


Figure 5.7. Validation of RNA-seq data upon DDX3X knock down in MCF7 cells.
(A) Table showing genes whose expression was down-regulated (*DDX3X, CCNA2, CDK2*), up-regulated (*KLF4*) or unchanged (*CCNE1*) in RNA-seq analysis of DDX3X knock down MCF7 cells. (B) mRNA levels of genes in A in MCF7 cells transfected with either scrambled siRNA or #6 or #8 DDX3X siRNAs, measured by qPCR. Cells were harvested 72h after transfection. Results represent the average of three biological replicates. (C) Western blot showing DDX3X, CDK2 and GAPDH protein levels in MCF7 cells used in B. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001; **** ≤ 0.0001).

Figure 5.8 Comparison of gene expression changes upon DDX3X knock down and DDX3X iCLIP targets in MCF7 cells.
Venn diagrams show the intersection between DDX3X iCLIP targets and genes up-regulated (left) or down-regulated (right) upon DDX3X knock down detected by RNA-seq.

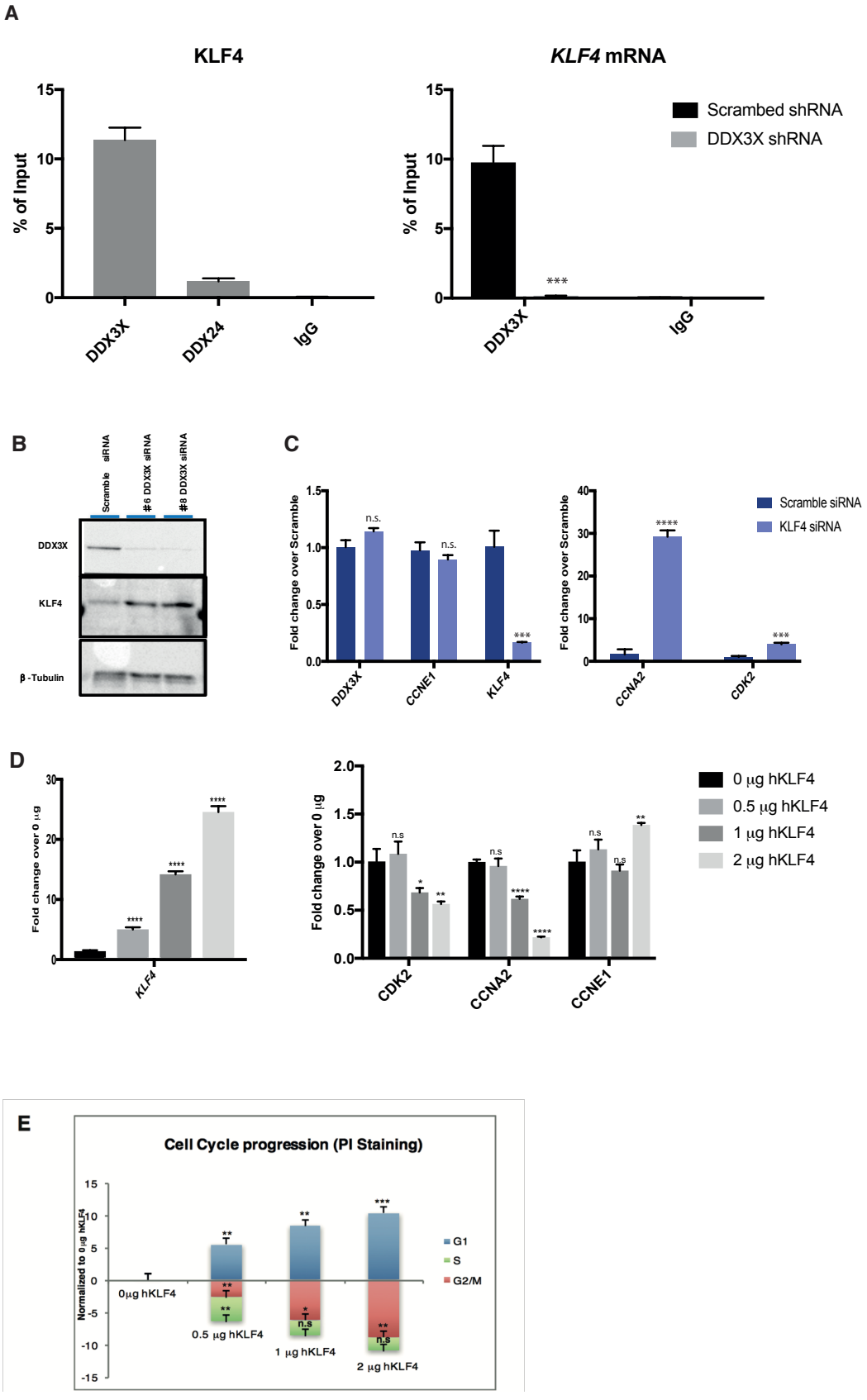

A
KLF4
KLF4 mRNA
% of Input
Scrambed shRNA
DDX3X shRNA
DDX3X
DDX24
IgG
***
B
Scramble siRNA
#6 DDX3X siRNA
#8 DDX3X siRNA
DDX3X
KLF4
β-Tubulin
C
Fold change over Scramble
n.s.
***
****
Scramble siRNA
KLF4 siRNA
DDX3X
CCNE1
KLF4
CCNA2
CDK2
D
Fold change over 0 μg
0 μg hKLF4
0.5 μg hKLF4
1 μg hKLF4
2 μg hKLF4
KLF4
CDK2
CCNA2
CCNE1
E
Cell Cycle progression (PI Staining)
Normalized to 0 μg hKLF4
G1
S
G2/M
0μg hKLF4
0.5 μg hKLF4
1 μg hKLF4
2 μg hKLF4

Figure 5.9. KLF4 mediates DDX3X-dependent regulation of cell cycle.
(A) CLIP-qPCR was performed as described in Methods (Section 2.13.3) in: (i) MCF7 cells, to validate binding of *KLF4* mRNA to DDX3X. DDX24 and IgG were used as controls (left panel). (ii) MCF7 cells transduced with either scrambled shRNA or DDX3X shRNA to validate the binding of *KLF4* mRNA to DDX3X. IgG was used as a control (right panel). cDNA obtained from each immunoprecipitation was analysed by qPCR to determine enrichment of the indicated targets. Results represent the average of three biological replicates. (B) Western blot shows the protein levels of KLF4, DDX3X and β-tubulin in MCF7 cells transfected with either scrambled siRNA or two different DDX3X targeting siRNAs (#6 or #8).
(C) mRNA levels of *DDX3X*, *CCNE1*, *KLF4*, *CCNA2* and *CDK2* in MCF7 cells transfected with either scrambled siRNA or *KLF4* targeting siRNA, measured by qPCR. Cells were harvested 72h after transfection. Results represent the average of three biological replicates.
(D) mRNA levels of *DDX3X*, *CCNE1*, *KLF4*, *CCNA2* and *CDK2* in MCF7 cells transfected with increasing amount of hKLF4 expressing plasmid as described in Methods (Section 2.8.3), measured by qPCR. Cells were harvested 48h after transfection. Results represent the average of three biological replicates.
(E) Cell cycle progression flow cytometric analysis in MCF7 cells transfected with increasing amount of hKLF4 expressing plasmid as in (D). Shown is the fold change quantification for each phase (G1-S-G2M). Cells were harvested 48h after transfection. Results represent the average of two biological replicates. Cells were treated with 50ng/ml nocodazole overnight.
Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to scramble (C) or 0μg controls (D-E) (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

5.6 Discussion

5.6.1 DDX3X and DDX54 show similar potential cell cycle regulation role

The results presented in this chapter highlight the role of DDX54 and DDX3X in regulating MCF7 cell growth through modulation of cell cycle. I found that both proteins are required for efficient proliferation of MCF7 cells and their progression through the cell cycle, specifically through S phase (Figure 5.4, Figure 5.5, Figure 5.6). This finding is completely new for DDX54, which has not been linked previously to cell cycle control. Conversely, DDX3X has previously been shown to regulate the cell cycle in different cell types (Section 1.5).

Interestingly, the gene ontology analysis of down-regulated genes in DDX54 and DDX3X knock down cells resulted in a very similar outcome. This suggests that the two proteins are involved in similar pathways and might have partly redundant function. However, when I compared the targets of DDX54 and DDX3X identified by iCLIP with the differentially expressed genes resulting from the knock down of the relevant helicase, I found virtually no overlap, apart from a very small number of candidates with no obvious common role (Figure 5.8 – Appendix Figure 4-5).

5.6.2 DDX3X is an upstream regulator of cell cycle

Notwithstanding the above, the intersection of DDX3X target mRNAs identified by iCLIP and down-regulated genes upon DDX3X knock down, highlighted one interesting candidate to follow up on: the transcription factor KLF4.

KLF4 is a zinc-finger containing transcription factor associated with growth arrest. In fact, the level of expression of KLF4 in actively proliferating cells is usually very low and it is lost in many different cancers [249-251]. As KLF4 possesses both a transactivation domain and a repression domain, its activity could have different outcomes based on the interaction with other transcriptional factors. The main function of KLF4

in the cell cycle is to induce arrest both at G1/S and G2/M transitions [248]. So far, this effect has been reported to be mediated, at least in part, by KLF4 dependent repression of *CCND1*, *CCNB1* and *CCNE1* [252].
KLF4 mRNA was validated as a direct target of DDX3X binding in CLIP-qPCR experiments and the levels of both *KLF4* mRNA and protein were strongly up-regulated upon DDX3X knockdown. These observations prompted me to investigate the link between DDX3X and KLF4 and to test the hypothesis that the cell cycle regulatory activity of DDX3X may be exerted through its regulation of KLF4 mRNA and protein levels.
In agreement with my hypothesis, knock down of KLF4 induced expression of *CCNA2* and *CDK2* genes that were down-regulated by DDX3X knock down. No change was observed in *CCNE1* levels, consistent with what was observed in DDX3X knock down cells. When *KLF4* was over-expressed, expression of *CCNA2* and *CDK2* was reduced proportionally to the level of expression of *KLF4* and consistent with the increased G1 arrest.
Association between cyclin A2 and CDK2 is necessary for progression through S phase and it is specifically required for DNA replication [253-256]. Interestingly, among the genes down-regulated upon DDX3X knock down in MCF7 cells, many genes encode replication-related factors such as the replication origin recognition complex 1 (ORC1), factors involved in DNA replication regulation (CDC6 and CDC45) and chromosome-related factors such as minichromosome maintenance proteins (MCM2, 3, 5, 7, 8,10) that associate with chromatin in G1 (raw data at link in Appendix Figure 3). This suggests DDX3X may have a broad role in regulating S phase-relevant factors.

DDX3X regulation of *CCNE1* mRNA translation is responsible for progression of HeLa cells to S phase [202]. In this system, DDX3X is required for production of CCNE1 protein but it does not affect *CCNE1* mRNA levels. Consistent with this, my data show that the mRNA level of *CCNE1* is not affected by DDX3X knock down in MCF7 cells. In the

previous chapter, I showed evidence for a potential role of DDX3X in positively regulating the levels of proteins translated from its bound mRNAs consistent with a positive translation regulatory activity of DDX3X (as previously proposed by others [203]).

In contrast, the findings presented in this chapter suggest a repressive role for DDX3X on the expression (at both the mRNA and protein level) of one of its target mRNAs, *KLF4*. It is possible that these represent distinct functions of DDX3X, which may be able to regulate different targets using different mechanisms.

5.6.3 CDKN1A (p21) involvement in DDX3X-dependent cell cycle regulation.

CDKN1A (p21) is a central regulator of cell cycle arrest in G1/S [257]. *P21* expression change was not detected in RNA-seq. Nevertheless, *p21* mRNA was found to be an iCLIP target of DDX3X and it harbours a CT (i.e. CU in RNA) rich binding motif in its 3'-UTR (CCTCCC –Appendix Figure 6). This observation potentially adds another level of complexity to DDX3X's regulation of the cell cycle. Certainly, further investigations are necessary as these might provide insight into a different mechanism through which DDX3X can counteract cell cycle arrest in MCF7 breast cancer cells.

Chapter 6 - Relevance of DDX54 and DDX3X in acute myeloid leukaemia cell lines

6.1 Introduction

The RNA-guided endonuclease CRISPR/Cas9 (clustered regularly inter-spaced short palindrome repeats/associated nuclease Cas9) system of bacterial host defence has recently revealed its powerful potential as a genome editing tool for the study of gene functions. In biomedical research, CRISPR/Cas9 represents a revolutionary approach to investigate disease-causing genes and to help identify new therapeutic targets.

Cas9 is an endonuclease that induces DNA double-strand breaks (DSB) whose activity can be directed to specific genomic sites by the sequence of a single guide RNA (sgRNA). Ectopic expression of Cas9 and a single guide RNA (sgRNA) in cells leads to formation of a DNA double-strand break (DSB) at specific genomic sites. In the absence of a homology-directed repair DNA template, DSBs are repaired by the non-homologous end-joining (NHEJ) pathway, which generates an assortment of short insertions and deletions (indels) in the target sequence. This can result in generation of a loss-of-function allele. Numerous studies have used CRISPR/Cas9 in mammalian cells that target 5' coding exons to create complete target gene knockouts [258, 259]. CRISPR/Cas9 can be used in both positive and negative selection screens. In the first case, it allows identification of tumour suppressors or genes that give growth advantage in particular conditions, such as drug resistance genes [258]. Negative selection or dropout screens allow identification of genes whose knockout confers lethality and determines consequent depletion of targeted cells [260]. Therefore, if applied in cancer cell lines, this approach can identify potential oncogenes. However, negative selection screens have shown limitations in their ability to uncover real oncogenes in cancer cells,

because of an enrichment of hits for essential genes such as RNA processing and DNA replication factors. Thus, performing negative screens in multiple cancer and non-cancer cell lines helps identify real oncogenes relevant for a particular cancer type. Another limitation of the common CRISPR/Cas9 approach is that it relies upon preferential targeting of 5' exons. Recently, Vakoc and colleagues reported that targeting the 5' exon of candidate genes often generates in-frame mutations that retain the functionality of the gene product thereby inhibiting identification of candidate oncogenes [261]. In the same work, the authors showed that CRISPR/Cas9 targeting of genomic regions encoding functional domains of a protein overcomes this limitation, since the resulting in-frame and frame-shift mutations within these domains results in loss of function alleles. Indeed, in this study the magnitude of negative selection obtained by targeting enzymatic domains was higher than that obtained when targeting 5′ exons or other regions. This indicates that a domain targeted CRISPR/Cas9 dropout approach can help identify functionally important domains within a target protein.

Results from recent CRISPR/Cas9 dropout screens, performed in various cancer cell lines, highlighted the importance of DDX54 and DDX3X in both mouse and human acute myeloid leukaemia (AML) cell lines [235]. As described in Section 1.6, *DDX3X* mutations have been identified in various blood cancers including T-cell acute lymphoblastic leukaemia, chronic lymphoblastic leukaemia and natural killer/T-cell lymphomas but nothing is known about the involvement of DDX3X in acute myeloid leukaemia (AML).

In this chapter, I present results from a CRISPR/Cas9 dropout approach in which the catalytic activities of DDX54 and DDX3X were targeted in mouse AML cells. Moreover, I describe results of a CRISPR/Cas9 dropout approach targeting distinct domains of DDX3X in OCI-AML2 and OCI-AML3 human AML cells.

6.2 Aims of this chapter

The aims of the work described in this chapter were:

- To determine the effect of targeting the helicase core of DDX54 and DDX3X in a mouse AML cell line (MLL-AF9) and in mouse NIH-3T3 fibroblasts using CRISPR/Cas9 technology
- To investigate the effect of targeting DDX3X functional domains in two humans AML cell lines (OCI-AML2 and OCI-AML3) using CRISPR/Cas9 technology

6.3 Targeting DDX54 and DDX3X catalytic domains affects MLL-AF9 mouse AML cell proliferation but not NIH-3T3 mouse fibroblast proliferation

In an attempt to link the function of RNA binders and modifiers to a cancer, the Kouzarides laboratory initiated a collaborative project with the Vakoc laboratory (Cold Spring Harbor Laboratories, USA) to perform a dropout screen in mouse AML cell lines using CRISPR/Cas9 technology. Previous work from the Vakoc laboratory showed that CRISPR targeting of gene regions encoding important functional domains of chromatin-associated factors significantly enhanced the efficacy of the dropout screens in identifying potential oncogenes [261].

In the CRISPR/Cas9 dropout screen we performed, mouse AML MLL-AF9/Cas9 cells, stably expressing Cas9, were infected with a pooled lentiviral sgRNA library to target the catalytic sites of different RNA-related enzymes. Five different sgRNAs were designed against each enzyme. Depletion of individual sgRNAs was measured by deep sequencing of the integrated sgRNA cassette in surviving cells over a 12-day time course. The read number of individual sgRNAs targeting the same gene were averaged at day 2 (first measurement point) and at day 12 (final measuring point) after infection. The ratio between average reads measured at day 2 and average reads measured at day 12 is called the

CRISPR score and represents one way of quantifying the magnitude of negative selection observed in a dropout screen.

Figure 6.1A graphically displays the CRISPR scores obtained for all the RNA hydrolases tested in the screen. In this study, a value of 50 was taken as the minimum of average reads considered at the measurement taken at day 2. Therefore, a value of CRISPR score above 50 indicates overgrowth of the targeted population whereas values of CRISPR score equal to or below 50 indicate no growth or depletion of the targeted population. To identify a biologically relevant range, a value of CRISPR score of 50 was sat as a maximum theoretical score and a value of CRISPR score of 10 was sat as a minimum significance threshold. The actual numerical CRISPR scores are listed in Figure 6.1B. Both *Ddx54* and *Ddx3x* scored significantly highly in the screen yielding CRISPR scores of 114.9 and 29.1, respectively (Figure 6.1A-B). These data highlight both *Ddx54* and *Ddx3x* as potential essential genes in mouse MLL-AF9 /Cas9 cells.

I validated the above results by performing the same CRISPR/Cas9 approach in both MLL-AF9/Cas9 cells and NIH-3T3/Cas9 cells. The AML cell line used in this study is the MLL-AF9/NRas^{G12D} AML cell line (RN2), which has been extensively used to identify genetic dependencies in leukaemia [262]. NIH-3T3 is an immortalized mouse fibroblast cell line that is used as a control. This cell line was used to discriminate between essential genes and genes specifically required for growth of AML cells. Genes showing depletion upon CRISPR/Cas9 catalytic targeting in NIH-3T3 cells are considered generally essential for cell survival and non-specific for AML cell growth. Cas9-expressing RN2 cells (MLL-AF9/Cas9 or RN2c) and NIH-3T3 cells (NIH-3T3/Cas9) were generated by the Vakoc laboratory using retroviral transduction with MSCV-hCas9-PGK-Puro, followed by puromycin selection and serial dilution to obtain single cell–derived clones. To validate results from the screen, I used two different sgRNAs designed to target the helicase core domain of *Ddx3x*, two sgRNA

against the helicase core domain of *Ddx54*, one sgRNA to target the negative control locus *Rosa26* and one sgRNA to target the essential (positive control) replication factor gene *Rpa3*. *Rosa26* is a genomic mouse locus commonly used for knock-in studies, whereas Rpa3 is an essential DNA replication factor whose targeting is lethal. The relevant sgRNA sequences were cloned into pLenti-U6-sgRNA-EFS-GFP vector for co-expression of both the sgRNAs and GFP. These plasmids were used to generate lentiviral particles bearing the relevant sgRNA and infect cells as described in Section 2.8.7.

Figure 6.2 shows the results of competition assays performed to measure negative selection. The rationale of a competition assay is that by performing infection at low viral titre, only a subset of cells will be targeted therefore leaving in culture a heterogeneous population of infected and non-infected cells. GFP fluorescent cells were monitored over time via flow cytometric analysis to measure out-competition of the GFP negative (non-infected) cells over the GFP positive (infected) cells. At each indicated time-point the percentage of GFP cells in the whole population was measured for all conditions.

The graphs in the left-hand panels of Figure 6.2 (panels A and C) show the ratio between the percentage of GFP positive cells observed at each time point over the percentage of GFP positive cells observed at the start (day 2 for MLL-AF9/Cas9 (panels A) and day 4 for NIH3T3/Cas9 (panels C)). GFP positive cells were monitored for 8 days. The fold depletion of GFP-positive cells over the whole period is the ratio of percentage of GFP positive cells observed at the starting time point over the percentage of GFP positive cells observed at the last time point. The fold depletion value was measured for both RN2/Cas9 (Figure 6.2B) and NIH3T3/Cas9 (Figure 6.2D) cells and is used to summarize the results of GFP progressive depletion (Figure 6.2A and C).

In MLL-AF9/Cas9 cells, targeting the region of the *Ddx54* gene encoding the catalytic domain led to strong depletion of GFP-positive cells with a

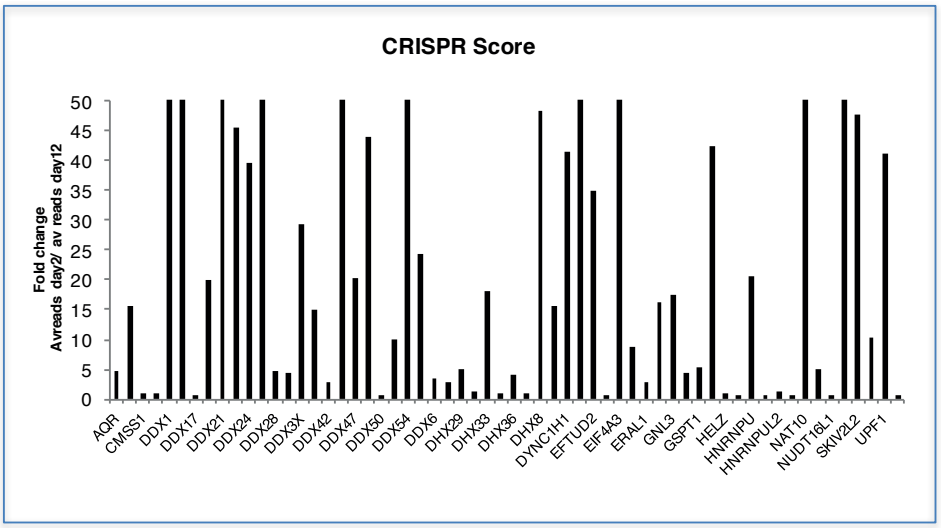
20- and 33-fold depletion for the two sgRNAs used (Figure 6.2B). Targeting of Ddx3x catalytic domain led to 13- and 15-fold depletions of GFP-positive cells. This is similar to what observed with *Rpa3* targeting that led to a 16-fold depletion. Targeting *Rosa26* did not cause depletion of GFP-positive cells.

In NIH-3T3/Cas9 cells (Figure 6.2C-D), no GFP-positive cell depletion was observed when targeting the gene regions encoding the catalytic domains of either *Ddx3x* or *Ddx54*. Targeting *Rosa26* caused negative depletion due to the GFP positive population increasing over time (Figure 6.2C). Targeting *Rpa3* caused a GFP-positive cell depletion of 2.8 fold (Figure 6.2D).

The difference in the magnitude of negative selection caused by targeting *Rpa3* between the two cell lines may be explained by their different growth rates. Indeed, MLL-AF9 cells express $NRas^{G12D}$ that confers growth advantage [262]. In any case, targeting *Rpa3* in NIH-3T3/Cas9 showed clear depletion of GFP positive cells after 8 days. In contrast, these cells showed no sensitivity to targeting of *Ddx54* or *Ddx3x*.

Overall, these observations show that growth sensitivity to CRISPR/Cas9 targeting of the regions of the *Ddx54* and *Ddx3x* genes encoding catalytic domains was not pan-lethal.

A



B

RNA HELICASE	CRISPR Score
DDX1	69.37
DDX10	92.4
DDX18	0.657
DDX21	19.94
DDX23	45.43
DDX24	39.53
DDX27	107.59
DDX28	4.63
DDX31	4.49
DDX3X	29.3
DDX41	14.85
DDX42	2.8
DDX46	88.82
DDX47	20.3
DDX49	43.93
DDX50	0.613
DDX51	9.97
DDX54	114.94
DDX56	24.35
DDX6	3.41
DHX15	2.93
DHX29	4.88
DHX30	1.17
DHX33	18.17
DHX34	0.821
DHX36	3.93
DHX57	0.958
DHX8	48.41
DHX9	15.73

Figure 6.1. CRISPR/Cas9 negative screening against RNA binders and modifiers highlights proteins potentially important in MLL-AF9 mouse cells survival.
(A) The graph shows the CRISPR score obtained from targeting different RNA hydrolases in a mouse MLL-AF9/Cas9 cell line. The CRISPR score equals the ratio of the average fold change of reads obtained at day 2 after infection over the average fold change of reads obtained at day 12 after infection for individual sgRNAs targeting the same gene. Fold change value of 50 threshold indicates the maximum theoretical score). Significance threshold was set at 5.
The screen was performed by Jungwei Shi (C. Vakoc lab) and Isaia Barbieri (T. Kouzarides lab). (B) The table shows the numerical CRISPR scores obtained in the screen and graphed in A. DDX3X and DDX54 showed significant CRISPR scores.

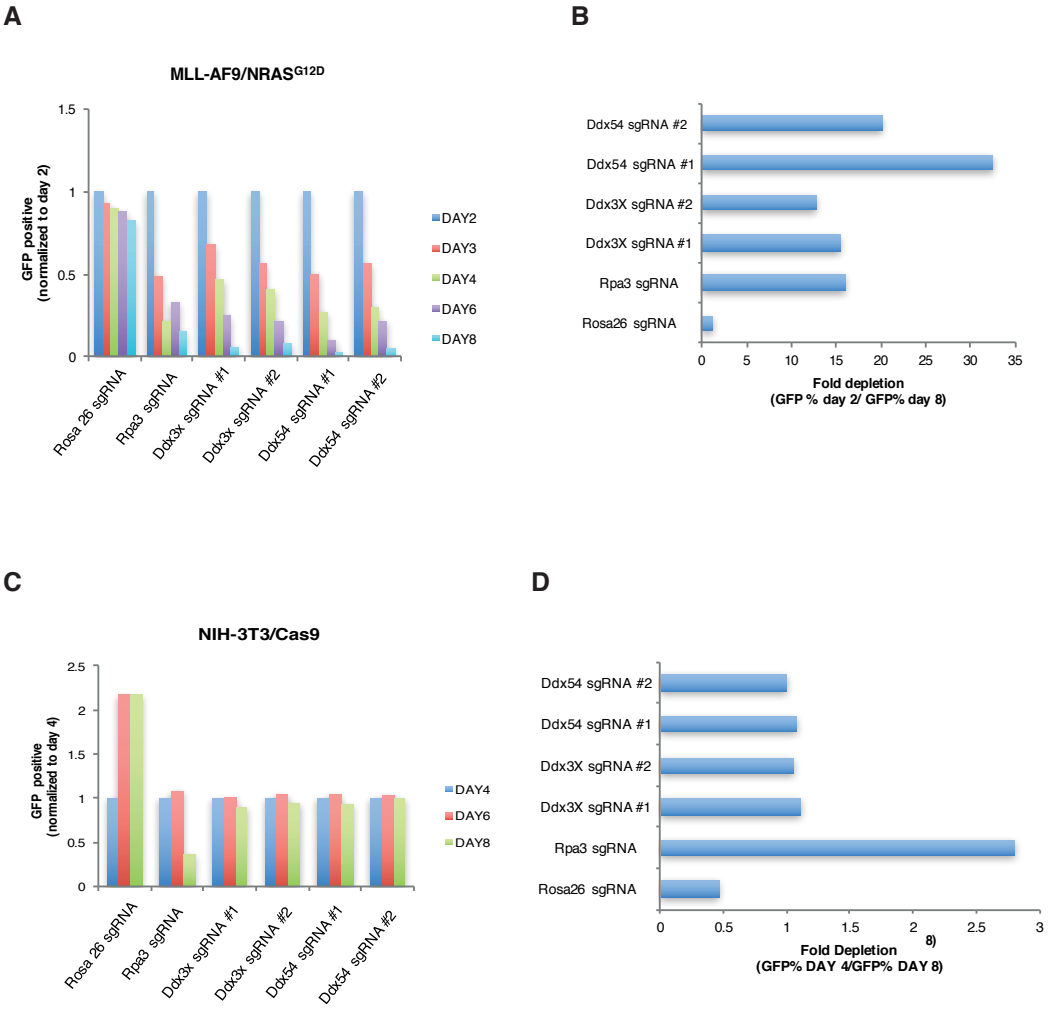


Figure 6.2. DDX54 and DDX3X targeting shows lethality in the AML MLL-AF9/Cas9 cell line but not in the control NIH-3T3/Cas9 cell line.
MLL-AF9/Cas9 (A-B) and NIH-3T3/Cas9 (C-D) cells were transduced with a lentivirus expressing *Rosa26, Ddx3x #1, #2, Ddx54 #1, #2* and *Rpa3* specific sgRNAs. Each sgRNA plasmid co-expresses a GFP marker.
(A-C). Ratio between the percentage of GFP positive cells at each time point and the percentage of GFP positive cells at the measurement start time point (day 2 for MLL-AF9/Cas9 (A) and day 4 for NIH-3T3/Cas9 (C)).
(B-D) Fold depletion of GFP-positive cells over the whole period, i.e. the ratio of percentage of GFP positive cells observed at the start (day 2 for MLL-AF9/Cas9 (B) and day 4 for NIH-3T3/Cas9 (D)) over the percentage of GFP positive cells observed at the last time point (day 8).

6.4 DDX3X functional domain targeting affects OCI-AML3 but not OCI-AML2 cells.

Very recently, a CRISPR/Cas9 dropout screen was published by Vassiliou and co-workers (Wellcome Trust Sanger Institute, Cambridge) [235]. The authors screened various human cancer cell lines, including those derived from colorectal adenocarcinoma (HT29), fibrosarcoma (HT1080) and blood cancers such as acute promyelocytic leukaemia (HL60), B myelomonocytic leukaemia (MV411) and acute myeloid leukaemia (AML: MOLM13, OCI-AML2, OCI-ALM3) [235]. In this work, the genes of several RNA related enzymes, including *DDX54* and *DDX3X*, were targeted at their 5′ exons in order to generate knock out cells. The results from the screen, from targeting *DDX54* and *DDX3X,* are summarized in Figure 6.3. Targeting *DDX54* resulted in significant dropout in all cancer cell lines tested, whilst targeting *DDX3X* resulted in significant dropout in only a subset of cancer cell lines. Amongst the leukaemia cell lines tested, acute myeloid leukaemia OCI-AML2 cells were resistant to *DDX3X* targeting whereas acute myeloid leukaemia OCI-AML3 cells were sensitive to loss of *DDX3X* [235].

Based on the above findings, I decided to further analyse the effects of targeting *DDX3X* in AML cells. Specifically, I selected the *DDX3X* knock out resistant OCI-AML2 cell line and the *DDX3X* knock out sensitive OCI-AML3 cell line for further analysis. I decided to use a CRISPR/Cas9 dropout approach to target regions of the *DDX3X* gene that encode different functional domains of the helicase. This approach was designed to provide insight into the importance of the different functional domains of this enzyme for growth of OCI-AML3 cells. In collaboration with the Vassiliou laboratory, I designed sgRNAs targeting region of the *DDX3X* locus encoding the different subdomains of DDX3X including the helicase core (RNA binding domain, ATP hydrolysis site and ATP binding loop), the N-terminal domain and the RS-like domain in the C-terminal region of DDX3X (Figure 6.4).

The Vassiliou laboratory generated OCI-AML2 and OCI-AML3 Cas9-expressing cell lines by lentiviral transduction using pKLV2-EF1aBsd2ACas9-W and selection in blasticidin (Section 2.8). sgRNAs targeting different regions of *DDX3X* encoding the different functional domains of DDX3X, as well as a positive control gene (*AURKB),* were cloned into pKLV2-U6gRNA5 (BbsI)-PKGpuro2ABFP-W vector bearing BFP marker and a puromycin resistance cassette. These plasmids were used to generate lentiviral particles bearing the relevant sgRNA or the empty vector and cells were infected as described in Section 2.8.7. Two different sgRNAs were used to target gene sequences encoding DDX3X's ATP hydrolysis site, ATP binding loop, N terminal domain and RS-like domain. Two sgRNAs were used to target the gene sequences encoding the two portions of the DDX3X RNA binding domain (referred to as "RNA binding domain A" and "RNA binding domain B") within the two helicase core domains. One sgRNA was used to target *DDX3X* exon 2 and one to target *DDX3X* exon 3. Two different sgRNAs were used to target the essential gene *AURKB* used as positive control of negative selection. AURKB is a key regulator of the mitotic spindle assembly, previously identified as essential in all cell lines tested in the published screen [235].
Competition assays were performed to measure negative selection comparing the results obtained with all sgRNAs in OCI-AML2/Cas9 and OCI-AML3/Cas9 cells (Figure 6.5). Under all conditions, cells were monitored over time via flow cytometric analysis to measure out-competition of the BFP negative (non-infected) cells over the BFP positive (infected) cells. At each indicated time-point the percentage of BFP cells in the whole population was measured for all conditions. BFP depletion is shown as the ratio of percentage of BFP positive cells observed at each time point over the percentage of BFP positive cells observed at the start (day 4) (Figure 6.5). BFP positive cells were monitored over 12 days. Fold depletion of BFP-positive cells over the 12 days period, was quantified as previously described for both OCI-AML3/Cas9 cells (Figure 6.6A) and OCI-

AML2/Cas9 cells (Figure 6.6B). The fold depletion values (Figure 6.6A-B) summarize the results of BFP progressive depletion (Figure 6.5).

In OCI-AML3/Cas9 cells, targeting of *DDX3X* 5′ exons 2 and 3 (which leads to DDX3X protein depletion - Figure 6.6C) led to an average 2 fold depletion of BFP-positive cells (Figure 6.6A). Targeting of the *DDX3X* gene sequences encoding the ATP hydrolysis site, ATP binding loop, N terminal domain or the RS-like domain, caused BFP-positive cells depletion from 1.4 to 1.8 fold. Interestingly, targeting the gene sequences encoding the two regions (A and B) of the DDX3X RNA binding domain led to higher depletion of BFP-positive cells ranging from 2.6 to 4.8 fold. When targeting the essential gene *AURKB,* BFP-positive cells depletion was 1.6 fold. Notably, this is below the value observed when the RNA binding domain of DDX3X was targeted (Figure 6.5, Figure 6.6A).

In OCI-AML2/Cas9 cells, most DDX3X targeting sgRNA, including the ones targeting DDX3X exons, caused no depletion of BFP-positive cells (Figure 6.6B). Targeting the essential *AURKB* gene generated an average BFP-positive cells depletion of 1.7. This is a higher fold depletion than any of those obtained by targeting DDX3X (Figure 6.6B). Interestingly, when targeting the RS-like C-terminal domain and ATP binding loop of DDX3X, the BFP-positive cells depletion fold change in OCI-AML2 reached the highest level observed for DDX3X targeting (~1.3 fold change). Infection with empty vector did not cause depletion of BPF-positive cells in either of the cell lines (Figure 6.5, Figure 6.6A-B).

These preliminary results confirm that DDX3X is selectively required for growth of OCI-AML3 but not OCI-AML2 leukaemia cells. Moreover, they indicate that DDX3X RNA binding function is more important than DDX3X enzymatic activity for OCI-AML3 cell survival.

Cancer-driver mutation :		MLLAF9	MLLAF4	NRAS	WT NPM1	MUT NPM1	KRAS	NRAS	MLLAF9	MLLAF4	NRAS
Cell line		MOLM13	MV411	HL60	OCI-AML2	OCI-AML3	HT29	HT1080	#AML	#non-AML	#All
		DDX54	DDX54	DDX54	DDX54	DDX54	DDX54	DDX54	5	2	7
				DDX3X		DDX3X	DDX3X		2	1	3

Dropout at FDR 20%
Adapted from Tzelepis et al; 2016

Figure 6.3. Sensitivity of human cancer cell lines to DDX3X or DDX54 targeting.
The table summarizes *DDX54* and *DDX3X* data from a CRISPR/Cas9 dropout screen in the indicated cell lines (Tzelepis et al. 2016). Dropout of the BFP-fluorescence positive population was set at a FDR of 20%. In each box, which corresponds to the indicated cancer cell line, the helicase(s) whose targeting gives significant depletion of the BFP-fluorescence positive population is indicated. Red box highlights result in OCI-AML2 and OCI-AML3 cells.
Solid tumour cells: colorectal adenocarcinoma (HT29), fibrosarcoma (HT1080); blood cancers: acute promyelocytic leukaemia (HL60), B myelomonocytic leukaemia (MV411) and acute myeloid leukaemia (MOLM13, OCI-AML2, OCI-ALM3). Top line lists major cancer driver mutations of each cell line.

A

Valiente-Echeverría et al., Review In Medical Virology (2015)

DDX3X DOMAIN	sgRNA target region amino acid coordinates
N-terminal domain	2-110
ATP binding loop	135-166
ATP binding and hydrolysis	195-232
RNA binding (on first helicase lobe) A	274-330
RNA binding (on second helicase lobe) B	445-506
RS-like (C-terminal)	582-632

Figure 6.4. Schematic representation of the CRISPR/Cas9 domain targeted approach against DDX3X.
Upper panel: Schematic structure of DDX3X indicating protein functional domains (adapted from [146]). Lower panel: Table shows amino acid regions of DDX3X targeted by the sgRNAs designed for CRISPR/Cas9 dropout approach.

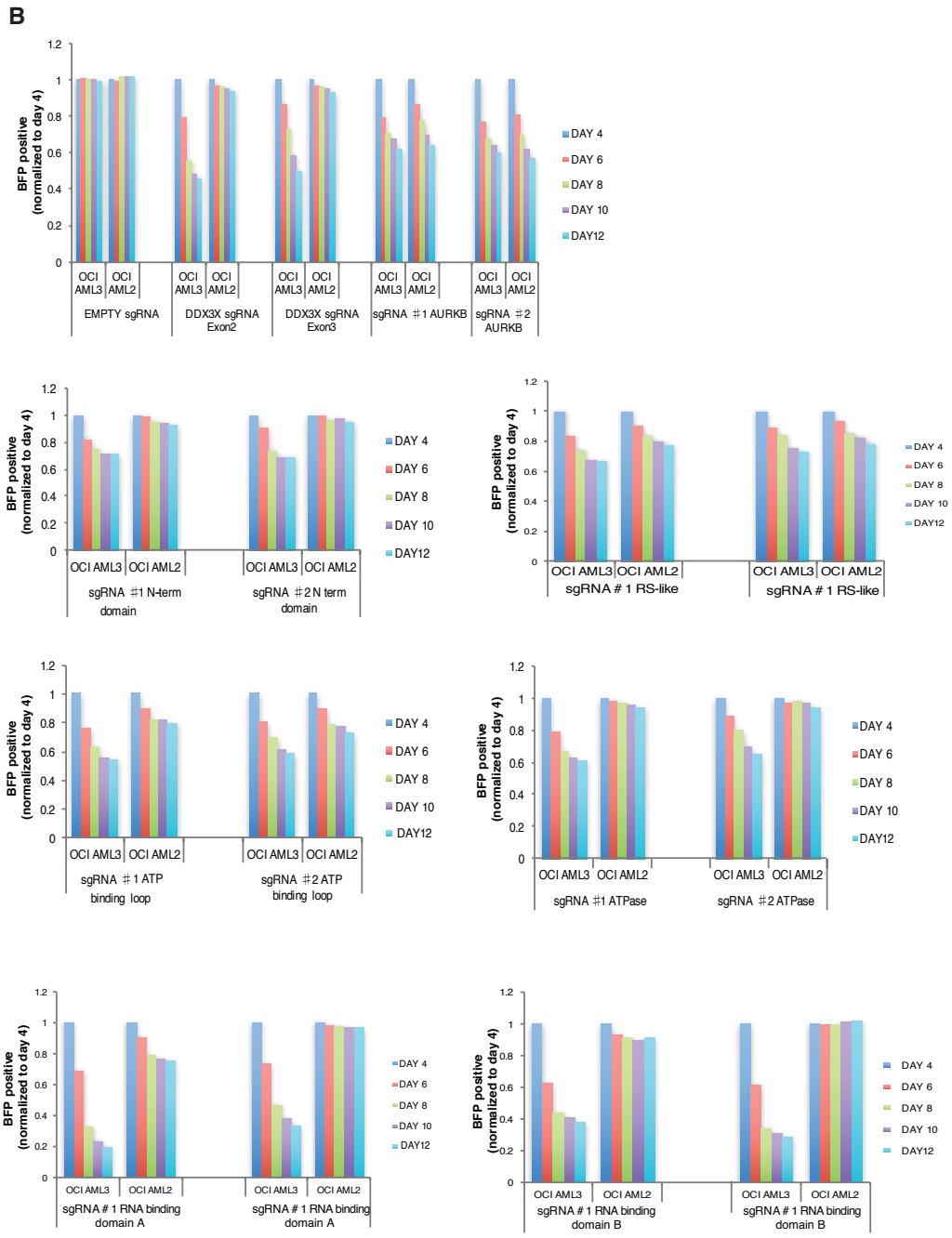


Figure 6.5. CRISPR/Cas9 targeting of functional domains of DDX3X in OCI-AML2/Cas9 and OCI-AML3/Cas9 AML cell lines: BFP-positive cell population depletion.

OCI-AML2/Cas9 and OCI-AML3/Cas9 cells were transduced with a lentivirus expressing *DDX3X* exon 2 and 3, N-terminal domain#1, #2, ATP binding loop#1, #2, ATP hydrolysis site #1, #2, RNA binding domain A#1, #2 and B #1, #2, RS-like C terminal domain #1, #2 and *AURKB* #1, #2 specific sgRNAs. Each sgRNA plasmid expresses a BFP marker and puromycin resistance cassette. All panels show the ratio between the percentage of BFP positive cells at each time point and the percentage of BFP positive cells at day 4 (the measurement starting time point) for both OCI-AML2/Cas9 and OCI-AML3/Cas9 cells infected with the indicated sgRNA construct or empty vector.

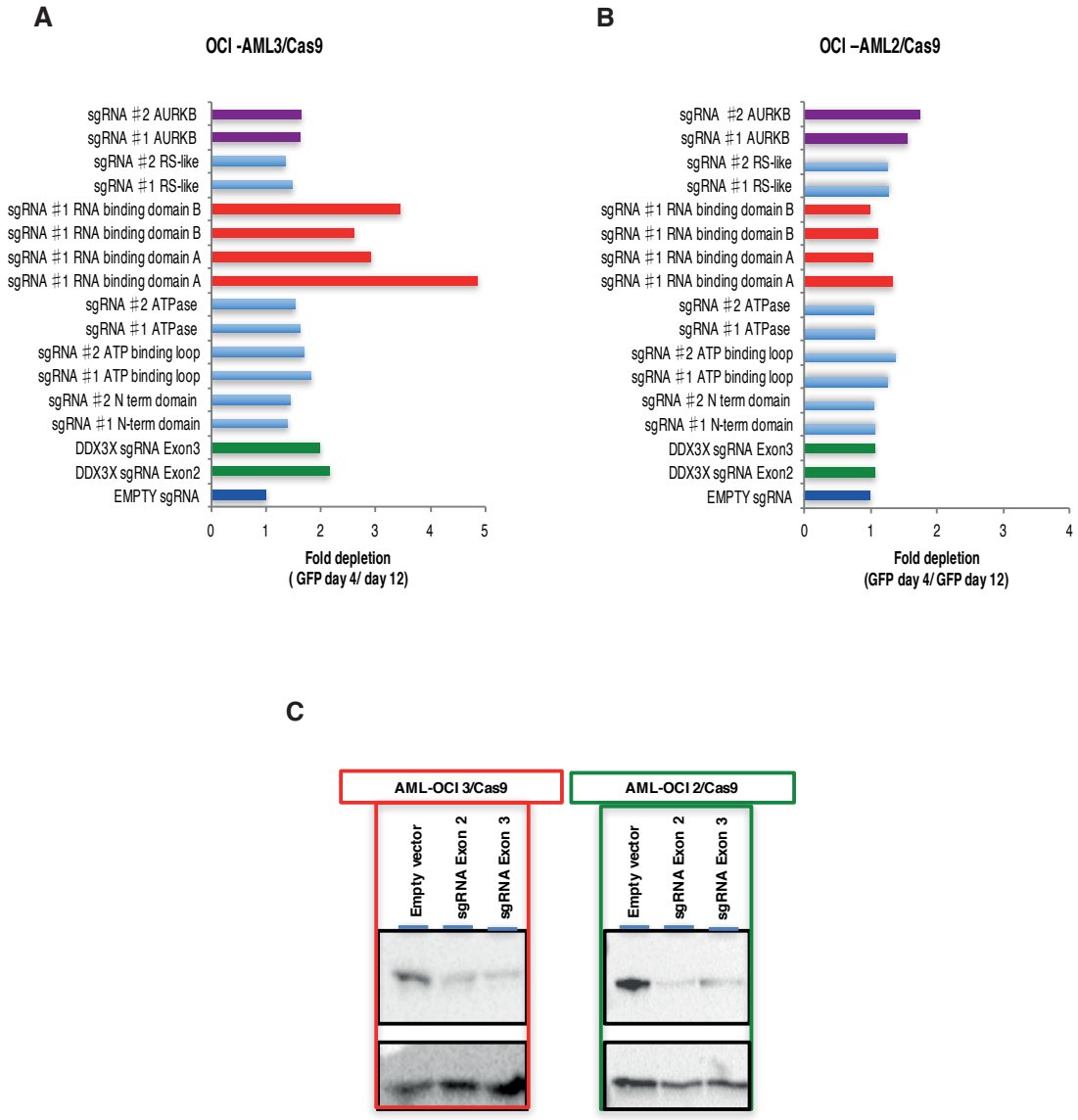


Figure 6.6. CRISPR/Cas9 targeting of functional domains of DDX3X in OCI-AML2/Cas9 and OCI-AML3/Cas9 AML cell lines: fold depletion.

(A-B) The panels show the fold depletion of BFP-positive cells over the entire duration of the experiment, i.e. the ratio of percentage of BFP positive cells observed at the starting time point (day4) over the percentage of BFP positive cells observed at the last time point (day 12) for both OCI-AML3/Cas9 (A) and OCI-AML2/Cas9 (B) cells infected with the indicated sgRNA construct or empty vector. (C) Western blot shows the protein levels of DDX3X and GAPDH in OCI-AML3/Cas9 cells (left panel) and OCI-AML2/Cas9 cells (right panel) infected with either empty vector, *DDX3X* exon 2 or *DDX3X* exon 3 specific sgRNA constructs for CRISPR/Cas9. Cells were harvested 4 days after selection in 1 μg/ml puromycin (6 days after infection).

6.5 Targeting the *DDX3X* gene sequence encoding the RNA binding domain reduces proliferation of OCI-AML3 cells and induces G1 cell cycle arrest

Based on the previous results from the domain targeted CRISPR/Cas9 approach against *DDX3X* in OCI-AML3/Cas9 and OCI-AML2/Cas9 cells, I next characterized the effect of targeting the gene sequences encoding the two regions of DDX3X's RNA binding domain on OCI-AML3/Cas9 and OCI-AML2/Cas9 cell proliferation. Each cell line was infected in triplicate with lentiviral particles bearing the relevant sgRNA, or the empty vector, and cells were monitored over time via flow cytometric analysis to measure out-competition of the BFP negative (non-infected) cells over the BFP positive (infected) cells. I used one sgRNA to target each of the two portions of DDX3X RNA binding domain (A and B) and one sgRNA to target the essential gene *AURKB*.

The results I previously obtained from targeting the same sequences were initially replicated in both cell lines using just one sgRNA per domain (Figure 6.7A-D). In fact, in OCI-AML3/Cas9 BFP-positive cell depletions were 3.5, 3.6 and 2 fold when targeting gene sequences encoding the two regions (A and B) of the DDX3X RNA binding domain and the *AURKB gene,* respectively (Figure 6.7 B). Again, no depletion of BFP positive cells was observed when targeting the same sequences of *DDX3X* in OCI-AML2/Cas9 cells, whereas a 1.6 fold depletion was obtained when targeting the *AURKB gene* (Figure 6.7 D).

Representative images from a flow cytometric analysis of targeted cells (as described in Figure 6.7) show the percentage of BFP-positive cells under all conditions at 4, 8 and 12 days after infection of OCI-AML3/Cas9 cells (Figure 6.8) and OCI-AML2/Cas9 cells (Figure 6.9). Targeting *DDX3X's* RNA binding domain encoding sequences induced a rapid decrease in the percentage of OCI-AML3/Cas9 BFP-positive cells over time, whereas the percentage of OCI-AML2/Cas9 BFP-positive cells remained constant.

Next, the OCI-AML3/Cas9 and OCI-AML2/Cas9 cells described above were selected in puromycin to isolate the BFP-positive population. Proliferation rates of these cells were then assessed (Figure 6.10A-B). Targeting DDX3X RNA binding domain (A and B) encoding sequences in OCI-AML3/Cas9 cells significantly reduced their proliferation to a level comparable to targeting the essential gene *AURKB* (Figure 6.10A). Specifically, at the last time point (day 6), the number of OCI-AML3/Cas9 cells in which RNA binding domain portions A and B had been targeted was on average about 35% less than the number of OCI-AML3/Cas9 cells infected with empty vector. *AURKB* targeted OCI-AML3/Cas9 cells grew about 30% less than OCI-AML3/Cas9 cells infected with empty vector, at the same time point (Figure 6.10A).

Conversely, OCI-AML2/Cas9 cells showed no reduced proliferation when DDX3X RNA binding domain (portions A or B) encoding sequences were targeted (Figure 6.10B). In contrast, when the *AURKB* gene was targeted, the number of cells was reduced by about 45% at the last time point (day 6) compared to OCI-AML2/Cas9 cells infected with empty vector (Figure 6.10B).

Since DDX3X knock down caused a cell cycle defect in MCF7 cells (Section 5.4), I performed a cell cycle analysis of OCI-AML2/Cas9 and OCI-AML3/Cas9 cells in which DDX3X RNA binding domain encoding sequences were targeted. BFP positive cell populations were selected in puromycin two days after infection, and grown in selection media for 6 additional days. Then, cells were treated overnight with 50 ng/ml nocodazole and tested the next day by propidium iodide staining in a flow cytometric analysis (Figure 6.11).

Targeting DDX3X RNA binding domains caused a dramatic increase of cells arrested in G1 phase in OCI-AML3/Cas9 cells compared to cells infected with empty vector (Figure 6.11A). Conversely, DDX3X RNA

binding domains targeted OCI-AML2/Cas9 cells did not show any consistent defect in cell cycle progression, compared to cells infected with empty vector (Figure 6.11A). Representative images of the propidium iodide staining-based flow cytometric analysis of data quantified in Figure 6.11 are shown in Figure 6.11B-C. When targeting DDX3X RNA binding domain (A and B) encoding sequences, the percentage of OCI-AML3/Cas9 cells in G1 phase was increased compared to cells infected with empty vector (Figure 6.11B). Conversely, the percentage of OCI-AML2/Cas9 cells in G1 phase remained unchanged compared to cells infected with empty vector (Figure 6.9C).

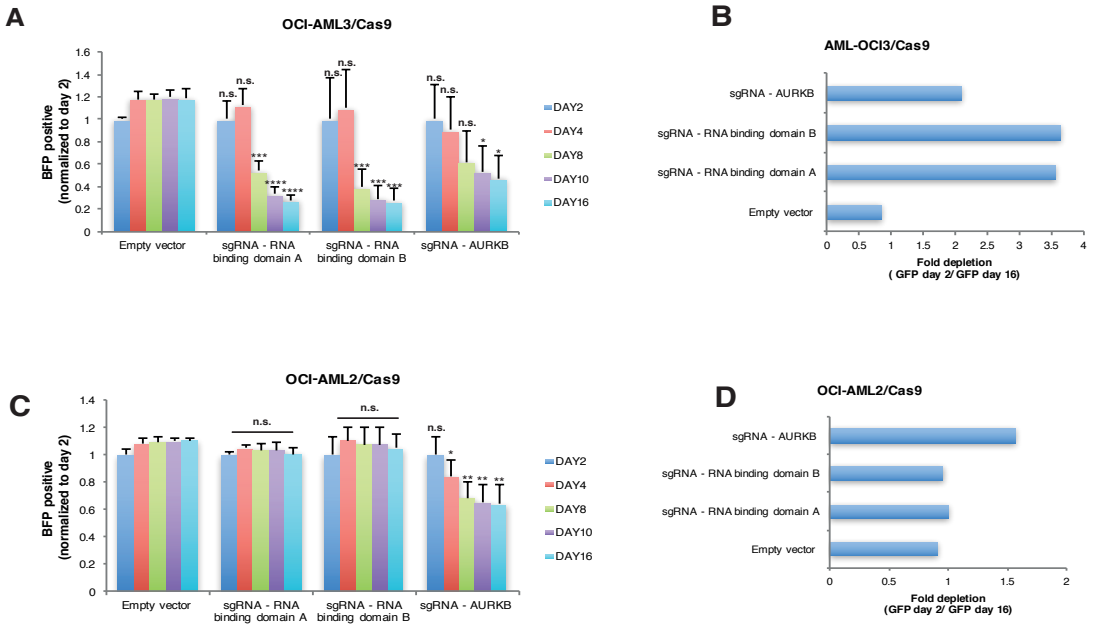


Figure 6.7. Validation of the effect of CRISPR/Cas9 targeting of DDX3X RNA binding domain in OCI-AML2/Cas9 and OCI-AML3/Cas9 cells.
(A and D) OCI-AML3/Cas9 (A and B) and OCI-AML2/Cas9 (C and D) cells were infected with *DDX3X* RNA binding domain A and B, *AURKB* specific sgRNAs or empty vector. (A-C) The ratio between the percentage of BFP positive cells at each time point and the percentage of BFP positive cells at day 2 (the measurement starting time point) was determined. (B and D) Fold depletion of BFP-positive cells over the entire experimental duration, i.e. the ratio of percentage of BFP positive cells observed at the starting time point (day 2) over the percentage of BFP positive cells observed at the last time point (day 16) is plotted. All conditions were produced in three biological replicates. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to empty control (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

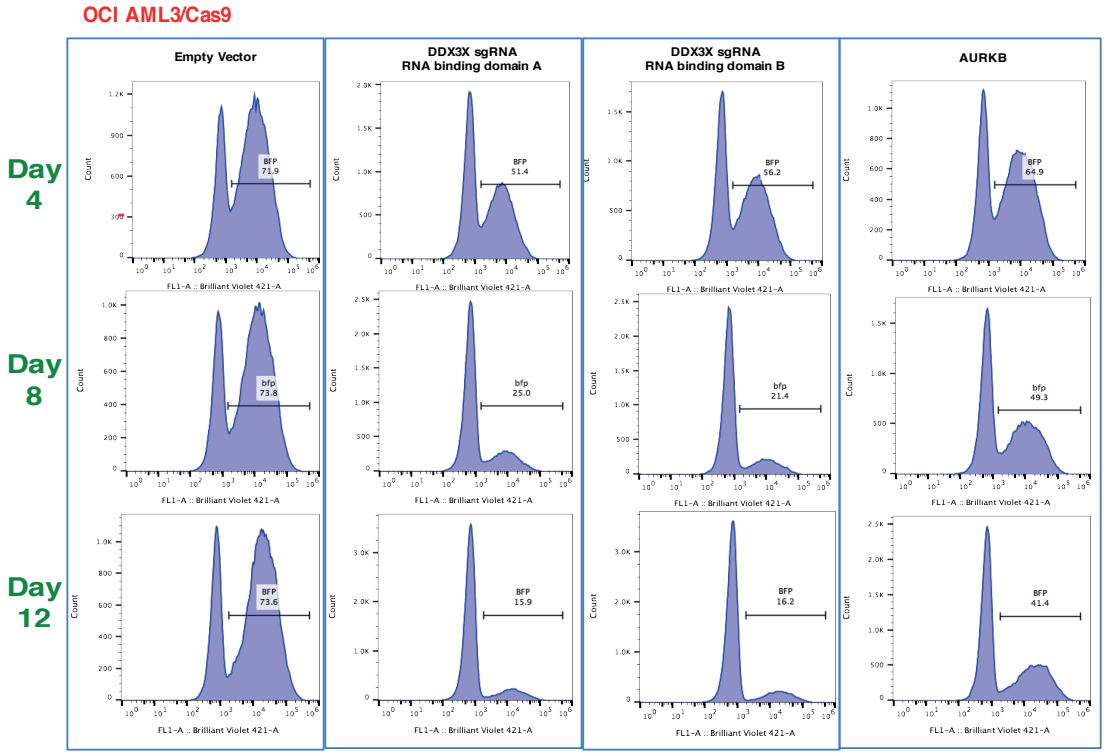


Figure 6.8. Flow Jo charts of OCI-AML3/Cas9 cells used in Figure 6.7A-B

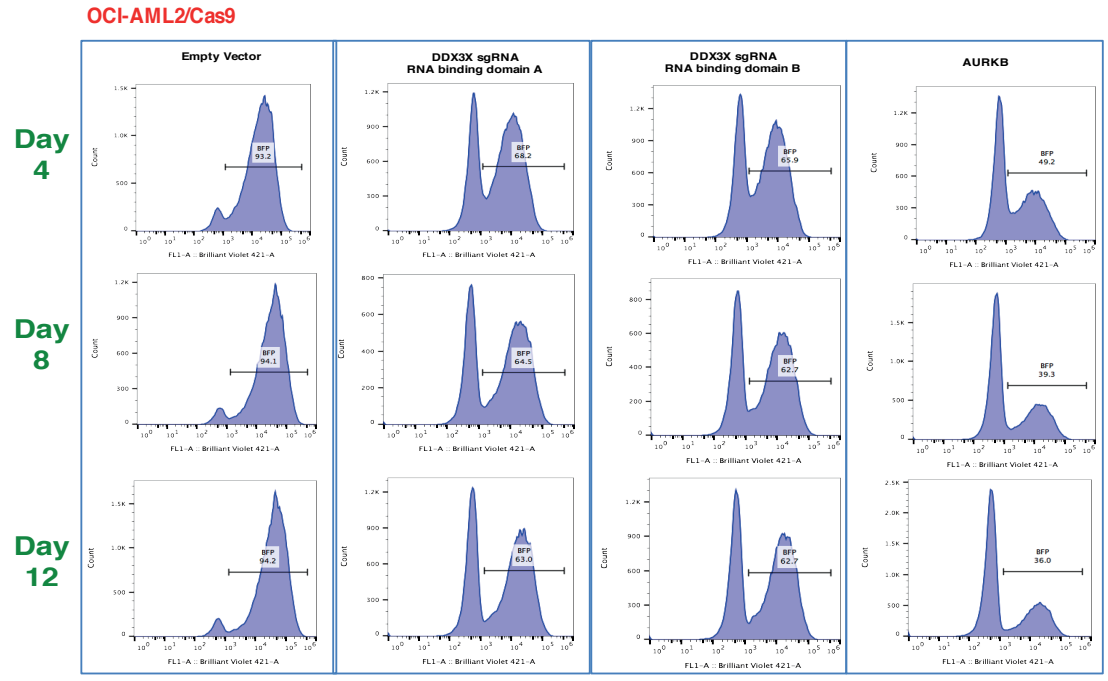


Figure 6.9. Flow Jo charts of OCI-AML2/Cas9 cells used in Figure 6.7C-D

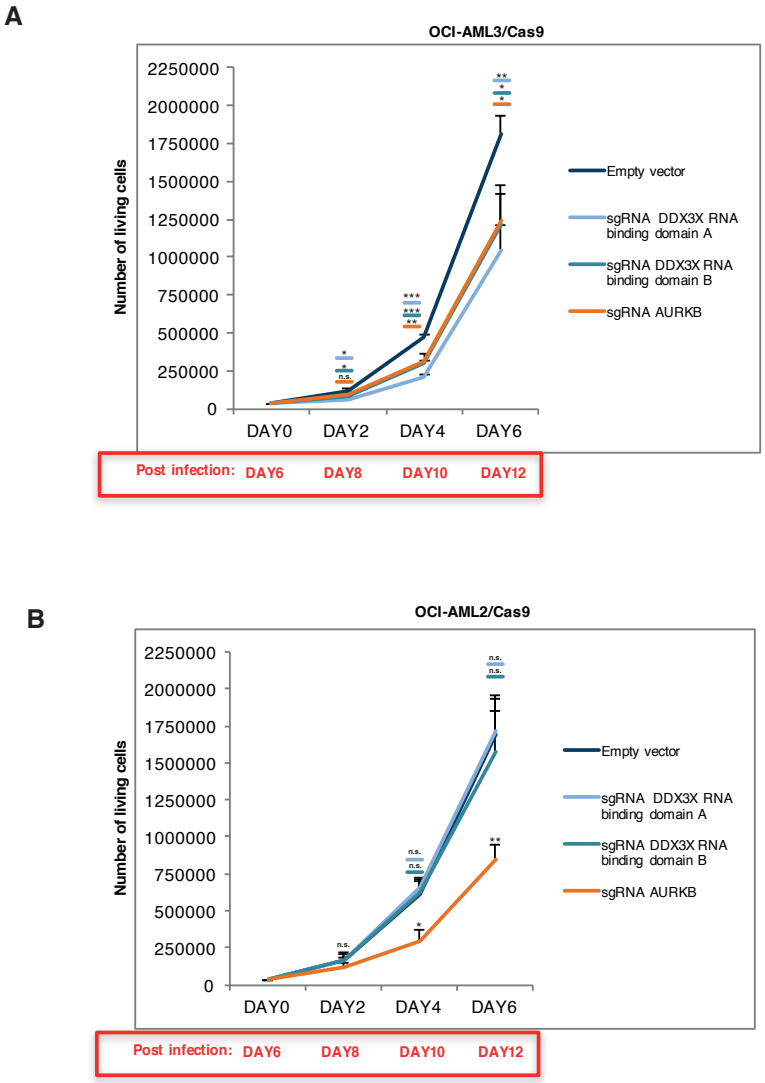


Figure 6.10. Disrupting DDX3X RNA binding domain reduces proliferation of OCI-AML3/Cas9 but not OCI-AML2/Cas9 cells.
Shown is the proliferation curve of OCI-AML3/Cas9 (A) and OCI-AML2/Cas9 (B) cells infected with DDX3X RNA binding domain A and B, AURKB specific sgRNAs or empty vector. Cells were selected in 1ug/ml puromycin from day 2 after infection for 4 days and then seeded for proliferation assay (see Section 2.8.2). Cells were counted at the indicated time-points after infection (red box). Result represents the average of three biological replicates. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to empty control (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001).

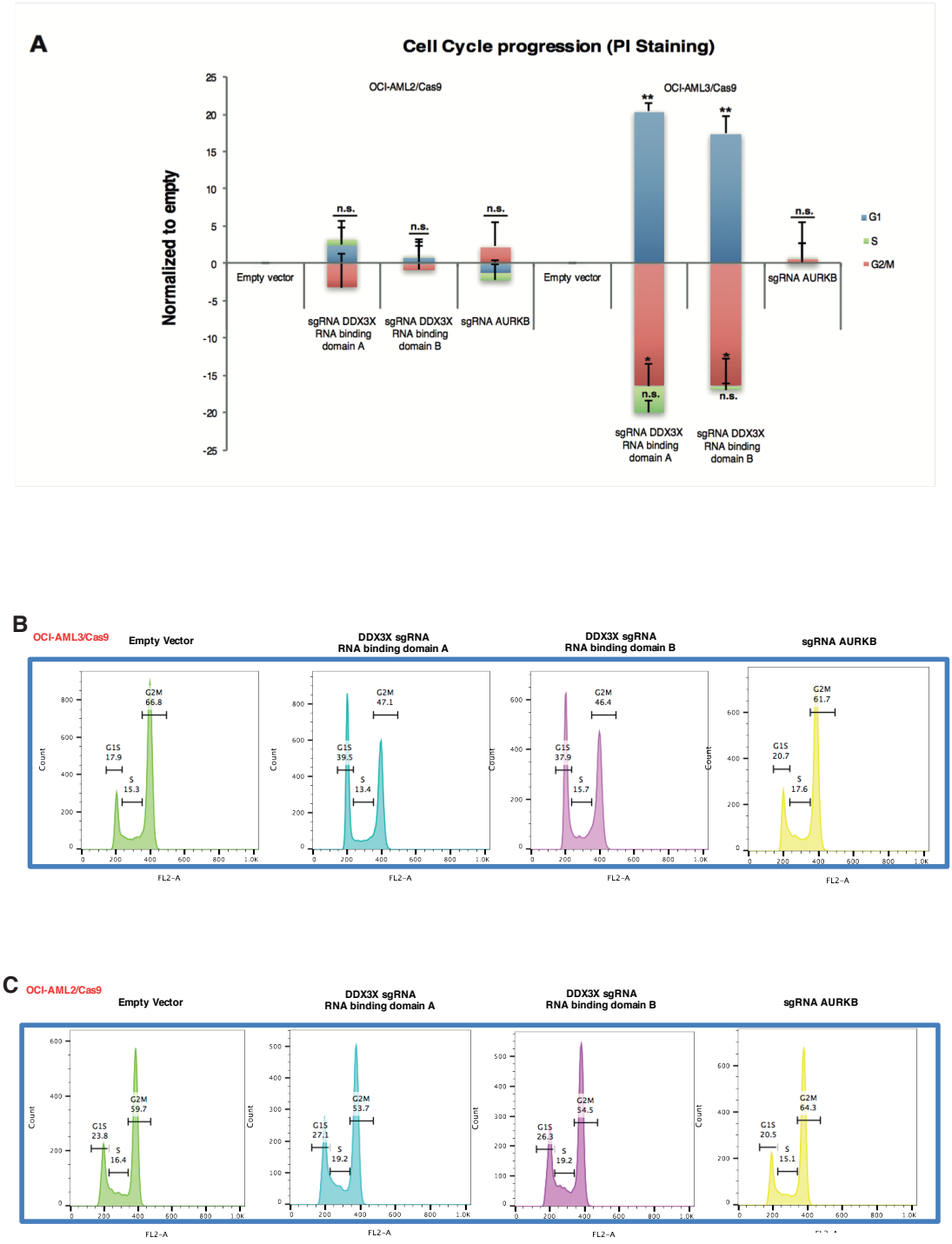


Figure 6.11. Disrupting DDX3X RNA binding domains results in G1 cell cycle arrest of OCI-AML3/Cas9 but not OCI-AML2/Cas9 cells.

Shown is the cell cycle progression flow cytometric analysis based on PI staining in OCI-AML3/Cas9 and OCI-AML2/Cas9 cells infected with DDX3X RNA binding domain A and B, *AURKB* specific sgRNAs or empty vector. Cells were selected in 1ug/ml puromycin from day 2 after infection for 6 days and then seeded for cell cycle analysis. Cells were treated overnight with 50ng/ml nocodazole overnight and harvested the next day for cell cycle analysis (day 8 after infection). (A) Shown is the fold change quantification per each phase (G1-S-G2M): values were normalized to empty vector value for both OCI-ALM3/Cas9 and OCI-ALM2/Cas9

cells. Results represent the average of three biological replicates. Error bars indicate standard deviation. P-values represent statistical significance calculated with unpaired t test compared to empty control (P-values: ns. > 0.05; * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001;**** ≤ 0.0001). (B-C) Flow Jo charts for OCI-ALM3/Cas9 (B) and OCI-AML2/Cas9 (C) cells referred to experiment in A.

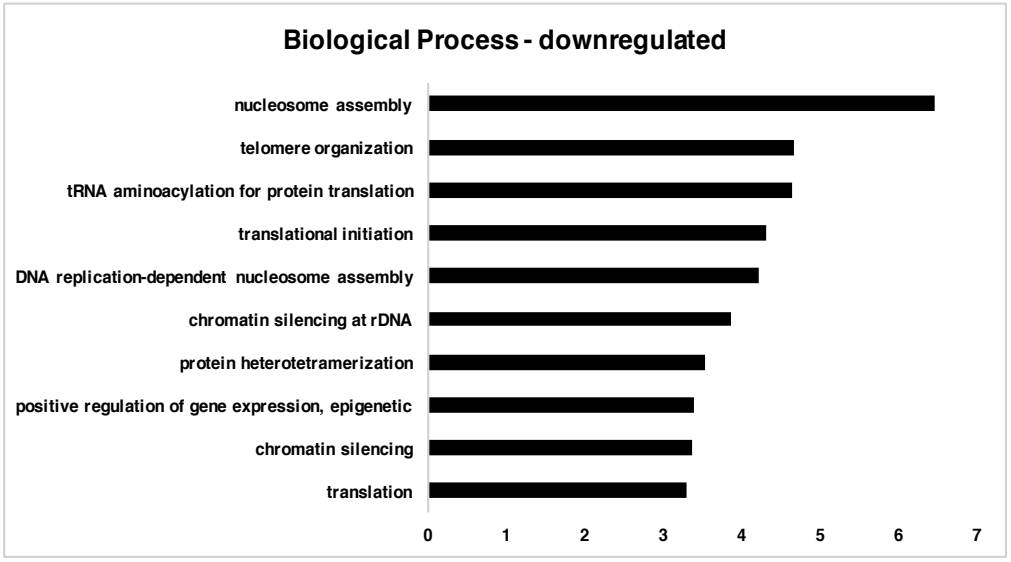
6.6 Gene expression changes upon disrupting DDX3X RNA binding domain in OCI-AML2 and OCI-AML3 cells

I next investigated the differential gene expression in OCI-AML2 and OCI-AML3 cells when the DDX3X RNA binding domain (A and B) was targeted by CRISPR/Cas9. Targeting of DDX3X RNA binding domain encoding sequences in OCI-AML2/Cas9 and OCI-AML3/Cas9 cells was performed as previously described (Section 6.5) and cells were harvested 8 days after infection (day 6 after selection in puromycin), which corresponded to the highest BFP depletion observed in OCI-AML3/Cas9 cells (Figure 6.7). RNA was isolated from cells and used for Illumina RNA-seq library preparation (Section 2.11- Appendix Figure 7A-B).

A DAVID GO term analysis was obtained with the common differentially expressed genes resulting from targeting both portions (A and B) of the DDX3X RNA binding domain in either cell lines. Gene ontology of down-regulated genes (Figure 6.12) shows enrichment for nucleosome and ribosome related terms in OCI-AML3/Cas9 cells. No significant GO terms for down-regulated genes were observed in OCI-AML2/Cas9 cells (Appendix Figure 7C), consistent with absence of phenotype shown by these cells when DDX3X's RNA binding domain was targeted (Figure 6.10, Figure 6.11).

Interestingly, similar terms were observed for down-regulated genes upon DDX3X knock down in MCF7 cells (Figure 5.3).

A



B

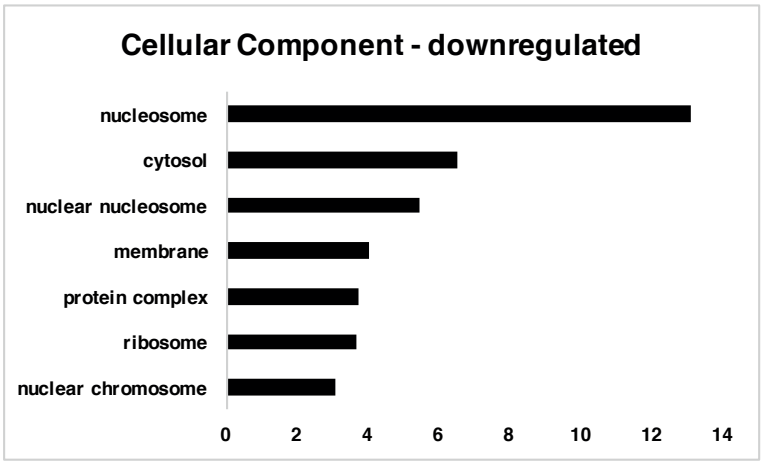


Figure 6.12. Gene expression changes in OCI-AML3/Cas9 cells upon targeting of the DDX3X RNA binding domain.

(A-B) Down-regulated genes in OCI-AML3/Cas9 cells upon targeting of DDX3X RNA binding domain were identified as described (Section 2.11.2) from RNAseq data (Section 2.11.1).

The GO enrichment of down-regulated genes was obtained using the DAVID functional annotation tool with default settings [PMID: 19131956]. The p-values were calculated using the "default" method of DAVID. X-axis shows the GO terms related to biological processes (A) or cellular components (B). Y-axis shows –log10 p-values of each GO term.

6.7 Discussion

The results of this chapter present evidence, for the first time, of a role for DDX54 and DDX3X in AML cells and they highlight a potentially critical role for DDX3X RNA binding activity in a particular subtype of AML, namely OCI-AML3 cells

6.7.1 AML cells are sensitive to the catalytic disruption of DDX54 and DDX3X

Although preliminary and requiring further replication, results from my validation of the two previously performed CRISPR/Cas9 dropout screens in both mouse and human cells, confirmed sensitivity of the leukaemia cell lines tested, to the catalytic impairment of the two DEAD box RNA helicases DDX54 and DDX3X. Specifically, catalytic targeting of both proteins resulted in a dramatic depletion of targeted mouse MLL-AF9/Cas9 cells, but this was not observed in targeted mouse NIH-3T3/Cas9 fibroblasts. These results indicate that the negative effect of targeting Ddx54 and Ddx3x is not due to a general lethality effect and that the function of these two proteins could be essential for maintenance of the leukaemia phenotype, whereas their functions may be dispensable in non-cancer cells, at least in mouse.

The data from the dropout screens performed by Tzelepis et al. [235] suggested a generally lethal effect of knocking out DDX54 in all cancer cell lines tested (Figure 6.3). This observation is to some extent consistent with what I observed in the mouse AML cells screen. In fact, although the targeting of this protein was not lethal in my validation of the screen in mouse NIH-3T3 fibroblasts, the CRISPR score obtained for DDX54 was considerably above the maximum theoretical score of 50 (Figure 6.1), beyond which CRIPSR score values indicate essential gene functions considered not specific to the cell line tested. Overall, these findings suggest that the role of DDX54 might be generally essential in many human cancer cell types.

In contrast to DDX54, targeting DDX3X revealed a selective sensitivity exhibited by only a subset of cancer cell lines. This prompted me to validate the effect of targeting different functional domains of this protein.

6.7.2 Targeting the DDX3X RNA binding domain impairs growth and cell cycle progression of OCI-AML3 but not OCI-AML2 AML cells

Evidence from the domain targeted CRISPR/Cas9 dropout screen I performed suggested that OCI-AML3 cells are specifically sensitive to loss of the RNA binding function of DDX3X more than loss of its enzymatic activity. This result, whilst confirming domain targeted CRISPR/Cas9 approaches as valid tools to select functionally important protein domains, further suggests that the ability of DDX3X to recognize specific RNAs may be more important for viability of specific cells than its ability to resolve double stranded structures within target RNAs. Furthermore, numerous studies have proposed that the hydrolysis activity of DDX proteins may be dispensable from their main biological function (Section 1.3). This would be consistent with my results obtained using sgRNAs versus DDX3X ATP binding loop and ATP hydrolysis site encoding sequences, which caused milder depletion of targeted OCI-AML3 cells.

The differences observed when results from the domain targeted screen in OCI-AML2 and OCI-AML3 cell lines are compared, may reflect the differential importance of the various functional domains of DDX3X in different cell lines. DDX3X may be part of specific protein complexes in OCI-AML3 cells distinct from the ones formed in OCI-AML2 cells, which could determine its RNA binding selectivity.

Importantly, targeting the RS-like C terminal domain of DDX3X, caused mild but similar sensitivity of the growth of both OCI-AML2 and OCI-AML3 cells. This motif is rich in arginine and serine residues often found in SR splicing factors (Section 1.2). Together with the observation that many DDX3X iCLIP target RNAs in MCF7 cells were found to contain a DDX3X binding motif with homology to the binding motif of splicing

factors SRSF1 and SRSF2 (Figure 4.11B), one speculation would be that DDX3X may have a splicing function that is required in all cells to some extent.

Impairment of DDX3X function in MCF7 cells through transient knock-down caused reduced proliferation and a G1 cell cycle arrest (Section 5.4). Importantly, a very similar phenotype was observed in OCI-AML3 cells when DDX3X RNA binding domain encoding regions were targeted by CRISPR/Cas9. This result suggests that DDX3X may function as a cell cycle regulator in different cancer cells and this role directly involves DDX3X RNA binding activity. Although I haven't investigated the binding specificity of DDX3X in OCI-AML3 cells, it seems plausible that inhibiting the function of DDX3X in different cancer cell lines, such as breast cancer cells and leukaemia cells, leads to cell cycle effects as a consequence of secondary effects due to DDX3X directly affecting stabilization or translation of mRNAs encoding cell cycle master regulators (such as the case of KLF4 in MCF7 cells, Chapter 4).

OCI-AML2 and OCI-AML3 cell lines are both AML cell lines with very different cytogenetic backgrounds. However, the main driver of the leukaemic phenotype in OCI-AML3 is known to be a type A mutation in the Nucleophosmin-1 gene (*NPM1*), which is not present in the OCI-AML2 cell line. This mutation consists in a TCTG insertion within exon 12 of the *NPM1* gene [263].

NPM1 is a predominantly nucleolar protein, which possesses structural domains allowing both cytoplasmic and nucleoplasmic localization. Under normal physiologic conditions, nuclear import predominates over export [264]. Translocation of NPM1 to the cytoplasm requires interaction of CRM1 and NPM1's nuclear export signal (NES) motifs [265]. Different functions of NPM1 have been described, including roles in ribosome biogenesis, maintenance of genomic stability and p53-dependent stress response [265].

Insertions in exon 12 of the *NPM1* gene are among the most common driver mutations in AML, found in 20–30% of cases, with type A being the most frequent insertion (~80% of cases on *NPM1* insertions)[266].
These mutations cause a frameshift in the C-terminus, which causes the generation of an additional NES motif. This results in stronger nuclear export than import, which eventually causes higher cytoplasmic localization of NPM1. The mutant form is therefore referred to as NPM1c. Although *NPM1* mutations are always heterozygous, NPM1c can still homo-dimerize with wild-type NPM1 and recruit it to the cytoplasm [266].
Different mechanisms have been suggested to account for NPM1c driving AML. Firstly, reduced nuclear NPM1 causes genomic instability due to lack of NPM1-dependent control of centrosome duplication. Secondly, mis-localization of NPM1c-interacting tumour suppressors (such as ARF) antagonizes cell-cycle arrest and apoptosis induction [264, 265]. However, transgenic mice studies indicate that NPM1c alone is insufficient to generate leukaemia, suggesting there may be a cooperative factor in NPM1c-driven AML [267].

Further studies are needed to investigate whether the presence of the *NPM1* mutation is a determinant for the sensitivity of the OCI-AML3 cells to DDX3X functional knockout. Interestingly, the DDX3X RNA binding domain targeting caused down-regulation of nucleosome and ribosome related genes. This gene ontology overlaps with the one found for DDX3X knock down-dependent down-regulated genes in MCF7 cells (Figure 5.2). Moreover, it recalls two of the main functions of Nucleophosmin-1 such as nucleosome remodelling, by its histone chaperone activity, and involvement in ribosome biogenesis.

Chapter 7 - Discussion

The aim of my thesis was to investigate the role of two members of the DDX RNA helicase family, DDX54 and DDX3X. Specifically, I set out to characterize the RNA binding properties of these proteins and to better understand their biological roles in two cancer model systems: MCF7 breast cancer cells and OCI-AML3 AML cells.

Very recently, DDX54 was reported to exert a splicing regulatory function in MCF7 cells exposed to DNA damage [183]. Upon exposure to ionising radiation, DDX54 relocates from nucleoli to the nucleoplasm where it binds pre-mRNAs. By performing PAR-CLIP, the authors reported that DDX54 binds mostly to intronic and exonic regions close to the 3′-ends of transcripts. Moreover, gene ontology analysis for DDX54-bound introns included chromosome organization and cell cycle factors. Under conditions of DNA damage, DDX54 was shown to interact with members of the spliceosome (including U2AF1), to be required for expression and splicing of its bound DNA damage-responsive RNAs and for survival of MCF7 cells [183].

Overall, the above observations are in good agreement with the findings from my DDX54 iCLIP. Interestingly, amongst the nucleotide 8-mers bound by DDX54, I identified enrichment of a motif homologous to the recognition sequence of U2AF2 and U2AF50, two other members of the spliceosome complex (Figure 3.8C). This motif is also present within *MTA3,* one of the validated iCLIP targets of DDX54. Here, the motif is located within an intron close to an exon/intron boundary (Appendix Figure 8A). This suggests that DDX54 may have a general function in regulating splicing, not only restricted to DNA damage conditions.

Unfortunately, I could not reproduce DDX54's co-repressive role in ERα-dependent transcription, previously observed in MCF7 cells [180]. My

approach involved measuring the expression of ERα–dependent genes upon knock down of DDX54 and in the presence or absence of estrogen, via qPCR analysis of the resulting transcripts. Although, no DDX54 dependent effect was observed, my methodology would not have measured post-transcriptional changes to the transcripts, such as alternative splicing. It remains possible therefore that DDX54 regulation of ERα-dependent transcripts occurs, if at all, at this level. Consistent with this possibility, I did detect an interaction between DDX54 and ERα via immunoprecipitation. Furthermore, my ChIP-seq analysis failed to detect co-localization of the two proteins on chromatin, suggesting they interact off chromatin, which is consistent with a post-transcriptional function of DDX54.

If DDX54 acts as a general splicing factor, it is reasonable to think that its function is essential for cell survival. In support of this idea, knock down of DDX54 was lethal in most cell lines tested in this project (MCF7, MCF10A and MLL-AF9 mouse leukaemia cells), as well as in many cancer cell types screened using a CRISPR/Cas9 dropout approach [235] (Figure 6.3). Although DDX54 catalytic knockout did not cause lethality in mouse fibroblasts, the results from Chapter 5 and those from Tzlepis et al, (Figure 6.3) strongly suggest the general essentiality of this protein at least in human cell lines. As shown, the biological function of DDX54 is ultimately due to its cell cycle regulation activity, at least in MCF7 breast cancer cells.

As with DDX3X, the intersection of DDX54 iCLIP targets and DDX54 knock down-induced gene expression changes, was extremely limited (Figure 5.8-Appendix Figure 5). This suggests that these two techniques provide very different information. Whereas direct biochemical function of RNA binding proteins may be inferred from iCLIP data, differential gene expression profiles likely highlight later, and therefore largely indirect effects due to depletion of the targeted protein.

One robust finding from my analyses is that DDX54 preferentially binds to introns. However, this is not in agreement with the strong association of DDX54 with *NEAT1* lncRNA, which is an intron-less transcript. Nevertheless, most of the DDX54 binding motifs found within *NEAT1* are AT/CT (i.e. AU/CU in RNA)-rich 8-mers (Appendix Figure 8B), which are similar to those found in intronic regions of other targets, which would suggest a sequence binding preference of DDX54.

NEAT1 lncRNA is up-regulated upon DNA damage and then forms additional paraspeckle structures [268]. Although my DDX54 iCLIP profile identified many binding sites within *NEAT1* lncRNA, and the binding was validated by CLIP-qPCR, I did not observe co-localization of DDX54 and *NEAT1* lncRNA in paraspeckles in MCF7 cells. Since both DDX54 and *NEAT1* lncRNA respond to DNA damage caused by different sources [183] [269], one explanation for this somewhat contradictory result may be linked to the fact that iCLIP and CLIP-qPCR both rely on UV irradiation of cells [270]. It is speculative but possible that the DNA damage response is triggered in cells exposed to UV light and that this induces both *NEAT1* lncRNA up-regulation and DDX54 relocation to the nucleoplasm, where the interaction occurs. Nevertheless, since it was shown that that exposure to UV radiation did not cause increase in H2AX phosphorylation in MCF7 cells and therefore does not induce DNA damage [183], this possibility is virtually unlikely.

Other speculations concerning the nuclear activity of DDX54 stem from its preferential binding to A-rich triplets and localization at the 3′ end of transcripts. DDX54 A-rich target sites are located in introns more than in 3'-UTRs. Therefore, DDX54 could potentially be involved in polyadenylation steps as well as in RNA editing steps, for instance in the unwinding of the dsRNA, coordinating the action of ADAR proteins (Section 1.1.1). As a future perspective, it would be interesting to investigate these hypotheses and to clarify the molecular mechanism underlying the cell cycle defect observed upon DDX54 knock down.

One striking result from my work was the strong overlap of genes down-regulated by both DDX54 and DDX3X knock down in MCF7 cells. This suggests that these two proteins function in very similar pathways. However, DDX54 is a nuclear enzyme whereas DDX3X is predominantly cytoplasmic in most cell lines tested, making it difficult to understand how their activities could be so redundant. It is possible that the nuclear fraction of DDX3X exerts much of its biological function. Nevertheless, the reason for such a strong overlap in common gene expression changes upon knockdown of either helicases remains elusive.

However, most published evidence so far, has linked DDX3X to cytoplasmic RNA processes (Section 1.5). Results from my studies tend to confirm this view (Chapter 4). I showed that DDX3X is required for maintaining the levels of proteins encoded by validated bound mRNAs (Figure 4.6) which could corroborate the hypothesis of a major cytoplasmic function in either counteracting degradation or favouring translation. In my results, DDX3X is enriched at both 5′ and 3′ ends of mRNAs (Figure 4.2). Taking into account the circularization of mRNAs occurring during translation by the interaction of poly-A tails- and cap- binding factors (Section 1.1.3), this observation could indicate a role for DDX3X in translation. In fact, DDX3X could be relevant for the bridging of 5′cap- and poly A tails- associated factors and therefore for the stability of mRNAs during translation.

A recently published CLIP profile identified binding of DDX3X to exonic regions of translation factors mRNAs, in agreement with my findings [203]. This work suggested a global role for DDX3X in regulating translation, mainly through recruitment of bound mRNAs into stress granules [203]. In contrast, my data show evidence for an effect of DDX3X activity on the levels of proteins encoded by specific mRNAs. Notably, DDX3X specifically binds to the *JUND* transcript and knockdown of the helicase leads to lower levels of JUND mRNA and protein. I identified a specific sequence within the *JUND* transcript to which DDX3X binds. This

was confirmed by EMSA experiments and the sequence represents the first validated DDX3X specific RNA binding sequence.
JUND is a member of the AP-1 family of dimeric transcription factors which can either activate or repress gene expression. Whereas other members of the AP-1 family have mainly oncogenic roles, JUND is generally associated with negative regulation of cell proliferation and positive regulation of cell differentiation [271]. Nevertheless, it was recently shown to be required for proliferation of prostate cancer cells [272]. The ultimate role of JUND seems to be highly dependent on the cell context as well as on the identity of its dimerization partner in the AP-1 dimer [273]

Interestingly, *JUND* expression has been shown to be regulated post transcriptionally: the binding of the DEAH-BOX helicase 9 (DHX9 or RHA) to *JUND* mRNA's 5'UTR is required for its efficient translation [247]. It will be interesting to investigate the biological significance of DDX3X binding to its specific recognition motif on the *JUND* mRNA sequence. In fact, it remains to be evaluated whether DDX3X affects JUND levels via its binding motif and whether the expression of JUND downstream DDX3X is relevant to cell survival.

Interestingly, the same motif found and validated in the *JUND* mRNA sequence was also found in *RPS2* mRNA. Amongst the validated ribosomal proteins (RPs) iCLIP targets, RPS2 displayed the highest sensitivity to DDX3X levels. In fact, RPS2 protein was almost completely depleted upon DDX3X knock down, whereas RPS6 and RPS3 protein levels showed little or no reduction, respectively. Notably, these RPs, displaying differential sensitivity to DDX3X depletion, harbour distinct DDX3X binding motifs within their sequences. Moreover, through structural prediction analysis, I found evidence supporting DDX3X binding to similar stem loop structures on *RPS2*, *RPS3* and *JUND* mRNAs. These results suggest that DDX3X binding to *RPS2, RPS3* and *JUND* mRNAs requires a region extended over the single strand and the

double strand portions of a stem loop. Although based on structural predictions, this finding represents a validation from *in vivo* data of the DDX proteins loading mechanism on RNA targets previously reported *in vitro* (Section 1.3). Moreover, it suggests that the structure of the binding site may be a direct element for recognition by DDX3X. Instead, the sequence of the binding motif may be more relevant for the functional consequences of DDX3X binding, perhaps by recruiting motif-specific effector proteins.
It is also possible that the mutation in the *JUND* probe tested in EMSA, disrupts the secondary structures necessary for DDX3X and DDX3X–containing complexes binding to the RNA *in vitro.*

Previous reports [96], as well as my iCLIP GO terms, have linked DDX3X activity to the mTOR pathway. Interestingly, I observed that the protein levels of RPS2, RPS3 and DDX3X itself were reduced upon rapamycin dependent mTOR inhibition. This suggests that rapamycin is acting on the translation level of these proteins in a DDX3X-independent way. Nevertheless, under this growth inhibitory condition, the binding of DDX3X to its target mRNAs remained relatively constant. One speculation about the functional role of DDX3X binding in conditions of growth arrest and translational inhibition such as the ones caused by rapamycin, is that the helicase may have a protective role against degradation of mRNAs that encode factors essential for cell growth.

The overall biological function of the binding features of DDX3X was not immediately evident from the sole analysis of iCLIP data, therefore I also evaluated gene expression as well as the cells growth parameters changes caused by depletion of DDX3X in cells. Knock-down of DDX3X in MCF7 cells caused a general reduction of proliferation due to a cell cycle arrest in G1. Interestingly, the same phenotype was also observed in OCI-AML3 leukaemia cells in which the RNA binding activity of DDX3X had been disrupted using a CRISPR/Cas9 approach. This result suggests that

DDX3X could have a general biological function which is not specific to one type of cancer. This was further corroborated by the fact that, in the absence of a functional DDX3X RNA binding domain in OCI-AML3 cells, many nucleosome-related genes were down-regulated, which was a gene expression changes signature very similar to the one observed in MCF7 cells depleted of DDX3X (Figure 5.4; Figure 6.12).

In MCF7 cells, the expression of cell cycle regulated genes involved in replication and onset of S phase, such as *CCNA2* and *CDK2* [254], were reduced by DDX3X knock down. These findings were surprising because none of them were identified in the DDX3X iCLIP dataset (Figure 5.8). On the other hand, gene ontology analysis of genes up-regulated by DDX3X knock-down mainly highlighted terms related to metabolic processes (Appendix Figure 4). Nevertheless, one of the very few targets resulting from the overlap of DDX3X iCLIP targets and up-regulated genes upon DDX3X knock down was the transcription factor and negative regulator of cell cycle, KLF4.

Further studies are needed to clarify the pathway of cell cycle regulation involving DDX3X and KLF4. The fact the *DDX3X* RNA level does not change upon KLF4 knock down, suggests DDX3X may be effectively upstream of KLF4 regulation in the proposed pathway. The hypothesised model therefore suggests that DDX3X is a key factor in repression of *KLF4* mRNA and protein, at least in cancer cells, thereby hindering transcriptional activation of key cell cycle genes (*CCNA2* and *CDK2*), which lie downstream of KLF4. Direct transcriptional regulation of KLF4 on these genes needs now to be confirmed. The mechanism of DDX3X-mediated repression of *KLF4* has to be clarified. DDX3X may act directly on the *KLF4* gene to repress its expression. However, this scenario is unlikely because of no evidence of DDX3X chromatin binding in my ChIP-seq data set (data not shown; raw data in Appendix Figure 2). Other possibilities concern the destabilization of the mRNA or induction of its degradation. DDX3X may even negatively regulate splicing of *KLF4* pre-

mRNA, which would account for a nuclear post-transcriptional function of DDX3X distinct from its main cytoplasmic roles reported so far.
The multifunctional mechanisms by which DDX3X regulates the protein levels of its target mRNAs may be explained, at least in part, by the fact that DDX3X exists in distinct macromolecular complexes (Section 1.5). Interestingly, although it differentially regulates the levels of proteins translated from its bound mRNAs, DDX3X activity is consistent with an oncogenic function: it favours cellular proliferation by stimulating ribosome production (via RPs), and it promotes cell cycle progression by impairing expression of cell cycle negative regulators, such as KLF4.

Overall, these observations corroborate the potential of DDX3X as a good candidate target for therapy development against certain cancers, although additional investigation is needed to understand its specific roles in different cell types. In this regard, a key novel finding of my thesis is the specificity of DDX3X oncogenic function in breast cancer cells compared to non-tumourigenic breast epithelial cells. This observation is particularly important when considering DDX3X as a target of onco-therapeutic intervention.

The essential role of DDX3X in AML cells is also a novel finding from my work. Indeed, the preliminary results obtained from studies of the two different leukaemia cells lines (OCI-AML2 and OCI-AML3) were particularly exciting. As with the breast cancer cell line data, these findings indicate a cell type-specific oncogenic role for DDX3X. Importantly, my work also identified a critical requirement for DDX3X's RNA binding activity, in addition to its ATPase activity, for sustained growth of leukaemic cells. This highlights DDX3X's RNA binding domains as potential therapeutically targetable regions, in addition to its enzymatic domain.

It remains to be determined why OCI-AML2 cells are insensitive to DDX3X functional knockout. Initially, it would be important to show that CRISPR targeting of the *DDX3X* locus occurred with equal efficiency in both cell lines. This will entail analysis of the CRISPR/Cas9 target sequences in the *DDX3X* locus in both cell lines in order to classify the percentage of indels induced by CRISPR/Cas9 and the number of in-frame or out-of-frame mutations occurring.

Since OCI-AML3 cells express a mutant version of the *NPM1* gene that determines aberrant localization of NPM1 in the cytoplasm (NPM1c), it would be interesting to investigate a potential NPM1c-related function of DDX3X in these cells that could then be responsible for their sensitivity to DDX3X functional targeting. For example, it may be that DDX3X cooperates with NPM1c to maintain a leukaemogenic phenotype in OCI-AML3. Indeed, this is particularly intriguing since mouse studies have shown that NPM1c is an AML-driver mutation but it is not sufficient to generate leukaemia on its own [267].

Since DDX3X has been shown to interact with CRM1 [193], which is the factor that mediates NPM1 and NPM1c export from the nucleus, one hypothesis is that DDX3X is required for the export of NPM1c, and therefore is necessary for NPM1c oncogenic activity in OCI-AML3. The notion of a DDX3X/NPM1 interaction and re-localization of NPM1 upon DDX3X knockout remains to be tested. It will also be essential to determine where DDX3X is localized within the two cell lines. In fact, DDX3X could mediate the potential oncogenic functions of NPM1c in the nucleus and/or the cytoplasm.

Overall, although leaving many open questions for experimental follow up, my results presented in this thesis have provided new insights into the mechanistic and biological functions of two members of the DDX family of RNA helicases. In particular, my observations highlight the relevance of DDX3X in the biology of a subset of cancer cells and underscore its importance as a potential therapeutic target.

Many DDX3 inhibitors have already been developed (Figure 7.1). These molecules are Ring expanded nucleoside (REN) analogues that mimic adenosine nucleosides and are designed to specifically target the ATP-binding cleft of DDX3 and impair the helicase activity [168].
Some of these compounds have been already tested for anticancer efficacy, such as the case of NZ-51. NZ-51 is a REN analogue that was found to inhibit viability and motility of breast cancer cells *in vitro*, but failed to show any efficacy *in vivo* [221].

A more recent and most studied DDX3 inhibitor is the REN analogue RK-33. RK-33 has shown specificity in binding DDX3 (whereas it did not bind other closely related DEAD-box RNA helicases) [216, 274] and to radio-sensitize lung and prostate cancer cells. Importantly, mice treatment with RK-33 did not cause normal tissue toxicity [216, 275].
Other DDX3 inhibitors have been produced and tested for antiviral therapy (specifically against HCV, HIV-1, West-Nile virus and dengue virus) [276-278] and could be evaluated for their anti-cancer potential.

Compound	Mechanism	Developmental status	Results
REN analogs (NZ51)	Target ATP-binding domain of DDX3	Tissue culture	Reduces cancer cell viability and motility
RK-33	Targets ATP-binding domain of DDX3	Preclinical animal studies	Causes radiosensitization in several mouse models and has single agent activity in a Ewing Sarcoma mouse model
NSC305787	Inhibitor of the DDX3-Ezrin interaction	Tissue culture	Anti-cancer activity not tested
Ketorolac salt (ZINC00011012)	Targets ATP-binding domain of DDX3	Preclinical animal studies	Anticancer activity in a mouse oral squamous cell carcinoma model
Rhodanine and Triazine derivatives	Target DDX3 ATP-binding domain	Tissue culture	In vitro anti-viral activity, anti-cancer efficacy not tested
16d	Inhibitor of RNA helicase activity	Tissue culture	in vitro anti-viral activity, anti-cancer efficacy not tested

Figure 7.1 DDX3 inhibitors.
The table summarizes the current knowledge about DDX3 inhibitors. Data were adapted from *Marise R. Heerma van Voss et al., 2017* [168]. (REN analogs NZ51 [221]; RK-33 [216, 218, 275]; NSC305787 [279]; Ketorolac salt [280]; Rhodanine and Triazine derivatives [277]; 16d [278]).

References

1. Glisovic, T., et al., *RNA-binding proteins and post-transcriptional gene regulation.* FEBS Lett, 2008. **582**(14): p. 1977-86.
2. Janga, S.C., *From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks.* Brief Funct Genomics, 2012. **11**(6): p. 505-21.
3. Cassar, P.A. and W.L. Stanford, *Integrating post-transcriptional regulation into the embryonic stem cell gene regulatory network.* J Cell Physiol, 2012. **227**(2): p. 439-49.
4. Lukong, K.E., et al., *RNA-binding proteins in human genetic disease.* Trends Genet, 2008. **24**(8): p. 416-25.
5. Kim, M.Y., J. Hur, and S. Jeong, *Emerging roles of RNA and RNA-binding protein network in cancer cells.* BMB Rep, 2009. **42**(3): p. 125-30.
6. Bentley, D.L., *Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors.* Curr Opin Cell Biol, 2005. **17**(3): p. 251-6.
7. Moore, M.J., *From birth to death: the complex lives of eukaryotic mRNAs.* Science, 2005. **309**(5740): p. 1514-8.
8. Stewart, M., *Nuclear export of mRNA.* Trends Biochem Sci, 2010. **35**(11): p. 609-17.
9. Marchese, D., et al., *Advances in the characterization of RNA-binding proteins.* Wiley Interdiscip Rev RNA, 2016. **7**(6): p. 793-810.
10. Bentley, D.L., *Coupling mRNA processing with transcription in time and space.* Nat Rev Genet, 2014. **15**(3): p. 163-75.
11. Beelman, C.A. and R. Parker, *Degradation of mRNA in eukaryotes.* Cell, 1995. **81**(2): p. 179-83.
12. Hocine, S., R.H. Singer, and D. Grunwald, *RNA processing and export.* Cold Spring Harb Perspect Biol, 2010. **2**(12): p. a000752.
13. Zipeto, M.A., et al., *RNA rewriting, recoding, and rewiring in human disease.* Trends Mol Med, 2015. **21**(9): p. 549-59.
14. Solomon, O., et al., *Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR).* RNA, 2013. **19**(5): p. 591-604.
15. Jiang, Q., et al., *ADAR1 promotes malignant progenitor reprogramming in chronic myeloid leukemia.* Proc Natl Acad Sci U S A, 2013. **110**(3): p. 1041-6.
16. Fumagalli, D., et al., *Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome.* Cell Rep, 2015. **13**(2): p. 277-89.
17. Qi, L., et al., *RNA editome imbalance in hepatocellular carcinoma.* Cancer Res, 2014. **74**(5): p. 1301-6.
18. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes Dev, 1997. **11**(21): p. 2755-66.
19. Zhao, J., L. Hyman, and C. Moore, *Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 405-45.

20. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 272-8.
21. Wahle, E. and U. Ruegsegger, *3'-End processing of pre-mRNA in eukaryotes.* FEMS Microbiol Rev, 1999. **23**(3): p. 277-95.
22. Drummond, D.R., J. Armstrong, and A. Colman, *The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in Xenopus oocytes.* Nucleic Acids Res, 1985. **13**(20): p. 7375-94.
23. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673-84.
24. Singh, P., et al., *Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes.* Cancer Res, 2009. **69**(24): p. 9422-30.
25. Masamha, C.P., et al., *CFIm25 links alternative polyadenylation to glioblastoma tumour suppression.* Nature, 2014. **510**(7505): p. 412-6.
26. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.
27. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome.* Nat Rev Mol Cell Biol, 2014. **15**(2): p. 108-21.
28. Dvinge, H., et al., *RNA splicing factors as oncoproteins and tumour suppressors.* Nat Rev Cancer, 2016. **16**(7): p. 413-30.
29. Fu, X.D. and M. Ares, Jr., *Context-dependent control of alternative splicing by RNA-binding proteins.* Nat Rev Genet, 2014. **15**(10): p. 689-701.
30. Jeong, S., *SR Proteins: Binders, Regulators, and Connectors of RNA.* Mol Cells, 2017. **40**(1): p. 1-9.
31. Han, J., et al., *SR proteins induce alternative exon skipping through their activities on the flanking constitutive exons.* Mol Cell Biol, 2011. **31**(4): p. 793-802.
32. Zahler, A.M., et al., *Distinct functions of SR proteins in alternative pre-mRNA splicing.* Science, 1993. **260**(5105): p. 219-22.
33. Maniatis, T. and B. Tasic, *Alternative pre-mRNA splicing and proteome expansion in metazoans.* Nature, 2002. **418**(6894): p. 236-43.
34. Hernandez-Lopez, H.R. and S.V. Graham, *Alternative splicing in human tumour viruses: a therapeutic target?* Biochem J, 2012. **445**(2): p. 145-56.
35. Wang, L., et al., *SF3B1 and other novel cancer genes in chronic lymphocytic leukemia.* N Engl J Med, 2011. **365**(26): p. 2497-506.
36. Quesada, V., et al., *Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia.* Nat Genet, 2011. **44**(1): p. 47-52.
37. Yoshida, K., et al., *Frequent pathway mutations of splicing machinery in myelodysplasia.* Nature, 2011. **478**(7367): p. 64-9.
38. Martin, M., et al., *Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3.* Nat Genet, 2013. **45**(8): p. 933-6.

39. Stephens, P.J., et al., *The landscape of cancer genes and mutational processes in breast cancer.* Nature, 2012. **486**(7403): p. 400-4.
40. Maguire, S.L., et al., *SF3B1 mutations constitute a novel therapeutic target in breast cancer.* J Pathol, 2015. **235**(4): p. 571-80.
41. Karni, R., et al., *The gene encoding the splicing factor SF2/ASF is a proto-oncogene.* Nat Struct Mol Biol, 2007. **14**(3): p. 185-93.
42. Anczukow, O., et al., *The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation.* Nat Struct Mol Biol, 2012. **19**(2): p. 220-8.
43. Cohen-Eliav, M., et al., *The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers.* J Pathol, 2013. **229**(4): p. 630-9.
44. Tange, T.O., et al., *Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core.* RNA, 2005. **11**(12): p. 1869-83.
45. Le Hir, H., et al., *The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions.* EMBO J, 2000. **19**(24): p. 6860-9.
46. Dostie, J. and G. Dreyfuss, *Translation is required to remove Y14 from mRNAs in the cytoplasm.* Curr Biol, 2002. **12**(13): p. 1060-7.
47. Le Hir, H., et al., *The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay.* EMBO J, 2001. **20**(17): p. 4987-97.
48. Nagy, E. and L.E. Maquat, *A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance.* Trends Biochem Sci, 1998. **23**(6): p. 198-9.
49. Kervestin, S. and A. Jacobson, *NMD: a multifaceted response to premature translational termination.* Nat Rev Mol Cell Biol, 2012. **13**(11): p. 700-12.
50. Long, J.C. and J.F. Caceres, *The SR protein family of splicing factors: master regulators of gene expression.* Biochem J, 2009. **417**(1): p. 15-27.
51. Singh, G., et al., *The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus.* Cell, 2012. **151**(4): p. 750-64.
52. Kohler, A. and E. Hurt, *Exporting RNA from the nucleus to the cytoplasm.* Nat Rev Mol Cell Biol, 2007. **8**(10): p. 761-73.
53. Lejeune, F., et al., *The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling.* EMBO J, 2002. **21**(13): p. 3536-45.
54. Okamura, M., H. Inose, and S. Masuda, *RNA Export through the NPC in Eukaryotes.* Genes (Basel), 2015. **6**(1): p. 124-49.
55. Viphakone, N., et al., *TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export.* Nat Commun, 2012. **3**: p. 1006.
56. Natalizio, B.J. and S.R. Wente, *Postage for the messenger: designating routes for nuclear mRNA export.* Trends Cell Biol, 2013. **23**(8): p. 365-73.
57. Culjkovic-Kraljacic, B. and K.L. Borden, *Aiding and abetting cancer: mRNA export and the nuclear pore.* Trends Cell Biol, 2013. **23**(7): p. 328-35.
58. Shen, A., et al., *Expression of CRM1 in human gliomas and its significance in p27 expression and clinical prognosis.* Neurosurgery, 2009. **65**(1): p. 153-9; discussion 159-60.
59. Huang, W.Y., et al., *Prognostic value of CRM1 in pancreas cancer.* Clin Invest Med, 2009. **32**(6): p. E315.

60. van der Watt, P.J., et al., *The Karyopherin proteins, Crm1 and Karyopherin beta1, are overexpressed in cervical cancer and are critical for cancer cell survival and proliferation.* Int J Cancer, 2009. **124**(8): p. 1829-40.
61. Dominguez-Sanchez, M.S., et al., *Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers.* BMC Cancer, 2011. **11**: p. 77.
62. Yang, J., et al., *Relationships of hHpr1/p84/Thoc1 expression to clinicopathologic characteristics and prognosis in non-small cell lung cancer.* Ann Clin Lab Sci, 2008. **38**(2): p. 105-12.
63. Guo, S., et al., *Linking transcriptional elongation and messenger RNA export to metastatic breast cancers.* Cancer Res, 2005. **65**(8): p. 3011-6.
64. Saito, Y., et al., *ALY as a potential contributor to metastasis in human oral squamous cell carcinoma.* J Cancer Res Clin Oncol, 2013. **139**(4): p. 585-94.
65. Lejeune, F. and L.E. Maquat, *Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells.* Curr Opin Cell Biol, 2005. **17**(3): p. 309-15.
66. Le Hir, H., J. Sauliere, and Z. Wang, *The exon junction complex as a node of post-transcriptional networks.* Nat Rev Mol Cell Biol, 2016. **17**(1): p. 41-54.
67. Fasken, M.B. and A.H. Corbett, *Process or perish: quality control in mRNA biogenesis.* Nat Struct Mol Biol, 2005. **12**(6): p. 482-8.
68. Lykke-Andersen, S. and T.H. Jensen, *Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes.* Nat Rev Mol Cell Biol, 2015. **16**(11): p. 665-77.
69. Gatfield, D. and E. Izaurralde, *Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila.* Nature, 2004. **429**(6991): p. 575-8.
70. Yamashita, A., et al., *Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover.* Nat Struct Mol Biol, 2005. **12**(12): p. 1054-63.
71. Cho, H., K.M. Kim, and Y.K. Kim, *Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex.* Mol Cell, 2009. **33**(1): p. 75-86.
72. Martin, L., et al., *Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense p53 mutations.* Cancer Res, 2014. **74**(11): p. 3104-13.
73. Medioni, C., K. Mowry, and F. Besse, *Principles and roles of mRNA localization in animal development.* Development, 2012. **139**(18): p. 3263-76.
74. Kapp, L.D. and J.R. Lorsch, *The molecular mechanics of eukaryotic translation.* Annu Rev Biochem, 2004. **73**: p. 657-704.
75. Aitken, C.E. and J.R. Lorsch, *A mechanistic overview of translation initiation in eukaryotes.* Nat Struct Mol Biol, 2012. **19**(6): p. 568-76.
76. Jorgensen, P. and M. Tyers, *How cells coordinate growth and division.* Curr Biol, 2004. **14**(23): p. R1014-27.
77. Laplante, M. and D.M. Sabatini, *mTOR signaling in growth control and disease.* Cell, 2012. **149**(2): p. 274-93.
78. Caron, E., et al., *A comprehensive map of the mTOR signaling network.* Mol Syst Biol, 2010. **6**: p. 453.

79. Oh, W.J. and E. Jacinto, *mTOR complex 2 signaling and functions.* Cell Cycle, 2011. **10**(14): p. 2305-16.
80. Mayer, C. and I. Grummt, *Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases.* Oncogene, 2006. **25**(48): p. 6384-91.
81. Roux, P.P. and I. Topisirovic, *Regulation of mRNA translation by signaling pathways.* Cold Spring Harb Perspect Biol, 2012. **4**(11).
82. Beretta, L., et al., *Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation.* EMBO J, 1996. **15**(3): p. 658-64.
83. Gingras, A.C., et al., *Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism.* Genes Dev, 1999. **13**(11): p. 1422-37.
84. Hara, K., et al., *Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism.* J Biol Chem, 1998. **273**(23): p. 14484-94.
85. Bhat, M., et al., *Targeting the translation machinery in cancer.* Nat Rev Drug Discov, 2015. **14**(4): p. 261-78.
86. Meyuhas, O., *Ribosomal Protein S6 Phosphorylation: Four Decades of Research.* Int Rev Cell Mol Biol, 2015. **320**: p. 41-73.
87. Wang, X., et al., *Regulation of elongation factor 2 kinase by p90(RSK1) and p70 S6 kinase.* EMBO J, 2001. **20**(16): p. 4370-9.
88. Jefferies, H.B., et al., *Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family.* Proc Natl Acad Sci U S A, 1994. **91**(10): p. 4441-5.
89. Levy, S., et al., *Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control.* Proc Natl Acad Sci U S A, 1991. **88**(8): p. 3319-23.
90. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.* Science, 2009. **324**(5924): p. 218-23.
91. Thoreen, C.C., et al., *A unifying model for mTORC1-mediated regulation of mRNA translation.* Nature, 2012. **485**(7396): p. 109-13.
92. Miloslavski, R., et al., *Oxygen sufficiency controls TOP mRNA translation via the TSC-Rheb-mTOR pathway in a 4E-BP-independent manner.* J Mol Cell Biol, 2014. **6**(3): p. 255-66.
93. Loreni, F., G. Thomas, and F. Amaldi, *Transcription inhibitors stimulate translation of 5' TOP mRNAs through activation of S6 kinase and the mTOR/FRAP signalling pathway.* Eur J Biochem, 2000. **267**(22): p. 6594-601.
94. Kawasome, H., et al., *Targeted disruption of p70(s6k) defines its role in protein synthesis and rapamycin sensitivity.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5033-8.
95. Tang, H., et al., *Amino acid-induced translation of TOP mRNAs is fully dependent on phosphatidylinositol 3-kinase-mediated signaling, is partially inhibited by rapamycin, and is independent of S6K1 and rpS6 phosphorylation.* Mol Cell Biol, 2001. **21**(24): p. 8671-83.

96. Tcherkezian, J., et al., *Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation.* Genes Dev, 2014. **28**(4): p. 357-71.
97. Hussey, G.S., et al., *Identification of an mRNP complex regulating tumorigenesis at the translational elongation step.* Mol Cell, 2011. **41**(4): p. 419-31.
98. Truitt, M.L. and D. Ruggero, *New frontiers in translational control of the cancer genome.* Nat Rev Cancer, 2016. **16**(5): p. 288-304.
99. Silvera, D., S.C. Formenti, and R.J. Schneider, *Translational control in cancer.* Nat Rev Cancer, 2010. **10**(4): p. 254-66.
100. Mamane, Y., et al., *eIF4E--from translation to transformation.* Oncogene, 2004. **23**(18): p. 3172-9.
101. Vezina, C., A. Kudelski, and S.N. Sehgal, *Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle.* J Antibiot (Tokyo), 1975. **28**(10): p. 721-6.
102. Heitman, J., N.R. Movva, and M.N. Hall, *Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast.* Science, 1991. **253**(5022): p. 905-9.
103. Hara, K., et al., *Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action.* Cell, 2002. **110**(2): p. 177-89.
104. Yang, H., et al., *mTOR kinase structure, mechanism and regulation.* Nature, 2013. **497**(7448): p. 217-23.
105. Oshiro, N., et al., *Dissociation of raptor from mTOR is a mechanism of rapamycin-induced inhibition of mTOR function.* Genes Cells, 2004. **9**(4): p. 359-66.
106. Janes, M.R. and D.A. Fruman, *Targeting TOR dependence in cancer.* Oncotarget, 2010. **1**(1): p. 69-76.
107. Benjamin, D., et al., *Rapamycin passes the torch: a new generation of mTOR inhibitors.* Nat Rev Drug Discov, 2011. **10**(11): p. 868-80.
108. Decker, C.J. and R. Parker, *P-bodies and stress granules: possible roles in the control of translation and mRNA degradation.* Cold Spring Harb Perspect Biol, 2012. **4**(9): p. a012286.
109. Protter, D.S. and R. Parker, *Principles and Properties of Stress Granules.* Trends Cell Biol, 2016. **26**(9): p. 668-79.
110. Wallace, E.W., et al., *Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress.* Cell, 2015. **162**(6): p. 1286-98.
111. Gilks, N., et al., *Stress granule assembly is mediated by prion-like aggregation of TIA-1.* Mol Biol Cell, 2004. **15**(12): p. 5383-98.
112. Kimball, S.R., et al., *Mammalian stress granules represent sites of accumulation of stalled translation initiation complexes.* Am J Physiol Cell Physiol, 2003. **284**(2): p. C273-84.
113. Tsai, N.P., P.C. Ho, and L.N. Wei, *Regulation of stress granule dynamics by Grb7 and FAK signalling pathway.* EMBO J, 2008. **27**(5): p. 715-26.
114. Buchan, J.R. and R. Parker, *Eukaryotic stress granules: the ins and outs of translation.* Mol Cell, 2009. **36**(6): p. 932-41.
115. Nover, L., K.D. Scharf, and D. Neumann, *Cytoplasmic heat shock granules are formed from precursor particles and are associated with a specific set of mRNAs.* Mol Cell Biol, 1989. **9**(3): p. 1298-308.

116. Liu, J., et al., *MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies.* Nat Cell Biol, 2005. **7**(7): p. 719-23.
117. Anderson, P. and N. Kedersha, *RNA granules: post-transcriptional and epigenetic modulators of gene expression.* Nat Rev Mol Cell Biol, 2009. **10**(6): p. 430-6.
118. Kedersha, N. and P. Anderson, *Stress granules: sites of mRNA triage that regulate mRNA stability and translatability.* Biochem Soc Trans, 2002. **30**(Pt 6): p. 963-9.
119. Ackerman, D. and M.C. Simon, *Hypoxia, lipids, and cancer: surviving the harsh tumor microenvironment.* Trends Cell Biol, 2014. **24**(8): p. 472-8.
120. Anderson, P., N. Kedersha, and P. Ivanov, *Stress granules, P-bodies and cancer.* Biochim Biophys Acta, 2015. **1849**(7): p. 861-70.
121. Arimoto, K., et al., *Formation of stress granules inhibits apoptosis by suppressing stress-responsive MAPK pathways.* Nat Cell Biol, 2008. **10**(11): p. 1324-32.
122. Kim, W.J., et al., *Sequestration of TRAF2 into stress granules interrupts tumor necrosis factor signaling under stress conditions.* Mol Cell Biol, 2005. **25**(6): p. 2450-62.
123. Moeller, B.J., et al., *Radiation activates HIF-1 to regulate vascular radiosensitivity in tumors: role of reoxygenation, free radicals, and stress granules.* Cancer Cell, 2004. **5**(5): p. 429-41.
124. Damgaard, C.K. and J. Lykke-Andersen, *Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR.* Genes Dev, 2011. **25**(19): p. 2057-68.
125. Ivanov, P., N. Kedersha, and P. Anderson, *Stress puts TIA on TOP.* Genes Dev, 2011. **25**(20): p. 2119-24.
126. Anantharaman, V., E.V. Koonin, and L. Aravind, *Comparative genomics and evolution of proteins involved in RNA metabolism.* Nucleic Acids Res, 2002. **30**(7): p. 1427-64.
127. Gerstberger, S., M. Hafner, and T. Tuschl, *A census of human RNA-binding proteins.* Nat Rev Genet, 2014. **15**(12): p. 829-45.
128. Lunde, B.M., C. Moore, and G. Varani, *RNA-binding proteins: modular design for efficient function.* Nat Rev Mol Cell Biol, 2007. **8**(6): p. 479-90.
129. Baltz, A.G., et al., *The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.* Mol Cell, 2012. **46**(5): p. 674-90.
130. Beckmann, B.M., A. Castello, and J. Medenbach, *The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions.* Pflugers Arch, 2016. **468**(6): p. 1029-40.
131. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.* Cell, 2012. **149**(6): p. 1393-406.
132. Shamoo, Y., N. Abdul-Manan, and K.R. Williams, *Multiple RNA binding domains (RBDs) just don't add up.* Nucleic Acids Res, 1995. **23**(5): p. 725-8.
133. Pereira, B., M. Billaud, and R. Almeida, *RNA-Binding Proteins in Cancer: Old Players and New Actors.* Trends Cancer, 2017. **3**(7): p. 506-528.
134. Bourgeois, C.F., F. Mortreux, and D. Auboeuf, *The multiple functions of RNA helicases as drivers and regulators of gene expression.* Nat Rev Mol Cell Biol, 2016. **17**(7): p. 426-38.

135. Jankowsky, E. and M.E. Fairman, *RNA helicases--one fold for many functions.* Curr Opin Struct Biol, 2007. **17**(3): p. 316-24.
136. Singleton, M.R., M.S. Dillingham, and D.B. Wigley, *Structure and mechanism of helicases and nucleic acid translocases.* Annu Rev Biochem, 2007. **76**: p. 23-50.
137. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters.* Curr Opin Struct Biol, 2010. **20**(3): p. 313-24.
138. Rogers, G.W., Jr., N.J. Richter, and W.C. Merrick, *Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A.* J Biol Chem, 1999. **274**(18): p. 12236-44.
139. Jankowsky, E., *RNA helicases at work: binding and rearranging.* Trends Biochem Sci, 2011. **36**(1): p. 19-29.
140. Rocak, S. and P. Linder, *DEAD-box proteins: the driving forces behind RNA metabolism.* Nat Rev Mol Cell Biol, 2004. **5**(3): p. 232-41.
141. Schutz, P., et al., *Comparative structural analysis of human DEAD-box RNA helicases.* PLoS One, 2010. **5**(9).
142. Tanner, N.K., et al., *The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis.* Mol Cell, 2003. **11**(1): p. 127-38.
143. Cordin, O., et al., *The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity.* EMBO J, 2004. **23**(13): p. 2478-87.
144. Bleichert, F. and S.J. Baserga, *The long unwinding road of RNA helicases.* Mol Cell, 2007. **27**(3): p. 339-52.
145. Linder, P. and E. Jankowsky, *From unwinding to clamping - the DEAD box RNA helicase family.* Nat Rev Mol Cell Biol, 2011. **12**(8): p. 505-16.
146. Soto-Rifo, R. and T. Ohlmann, *The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism.* Wiley Interdiscip Rev RNA, 2013. **4**(4): p. 369-85.
147. Liu, F., A. Putnam, and E. Jankowsky, *ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20209-14.
148. Yang, Q., et al., *DEAD-box proteins unwind duplexes by local strand separation.* Mol Cell, 2007. **28**(2): p. 253-63.
149. Chen, J.Y., et al., *Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor.* Mol Cell, 2001. **7**(1): p. 227-32.
150. de la Cruz, J., D. Kressler, and P. Linder, *Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families.* Trends Biochem Sci, 1999. **24**(5): p. 192-8.
151. Boudet, N., et al., *Evolution of intron/exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis, and Drosophila.* Genome Res, 2001. **11**(12): p. 2101-14.
152. Abdelhaleem, M., L. Maltais, and H. Wain, *The human DDX and DHX gene families of putative RNA helicases.* Genomics, 2003. **81**(6): p. 618-22.
153. Bowers, H.A., et al., *Discriminatory RNP remodeling by the DEAD-box protein DED1.* RNA, 2006. **12**(5): p. 903-12.

154. Cai, W., et al., *Wanted DEAD/H or Alive: Helicases Winding Up in Cancers.* J Natl Cancer Inst, 2017. **109**(6).
155. Fuller-Pace, F.V., *DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation.* Nucleic Acids Res, 2006. **34**(15): p. 4206-15.
156. Ilagan, J.O., et al., *Rearrangements within human spliceosomes captured after exon ligation.* RNA, 2013. **19**(3): p. 400-12.
157. Will, C.L., et al., *Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein.* EMBO J, 2002. **21**(18): p. 4978-88.
158. Kressler, D., et al., *Fal1p is an essential DEAD-box protein involved in 40S-ribosomal-subunit biogenesis in Saccharomyces cerevisiae.* Mol Cell Biol, 1997. **17**(12): p. 7283-94.
159. Yamazaki, T., et al., *The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression.* Mol Biol Cell, 2010. **21**(16): p. 2953-65.
160. Giorgi, C., et al., *The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression.* Cell, 2007. **130**(1): p. 179-91.
161. Gross, T., et al., *The DEAD-box RNA helicase Dbp5 functions in translation termination.* Science, 2007. **315**(5812): p. 646-9.
162. Fischer, N. and K. Weis, *The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1.* EMBO J, 2002. **21**(11): p. 2788-97.
163. Cordin, O., et al., *The DEAD-box protein family of RNA helicases.* Gene, 2006. **367**: p. 17-37.
164. Diges, C.M. and O.C. Uhlenbeck, *Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA.* EMBO J, 2001. **20**(19): p. 5503-12.
165. Nicol, S.M. and F.V. Fuller-Pace, *The "DEAD box" protein DbpA interacts specifically with the peptidyltransferase center in 23S rRNA.* Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11681-5.
166. Modelska, A., et al., *The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape.* Cell Death Dis, 2015. **6**: p. e1603.
167. Cencic, R. and J. Pelletier, *Throwing a monkey wrench in the motor: targeting DExH/D box proteins with small molecule inhibitors.* Biochim Biophys Acta, 2013. **1829**(8): p. 894-903.
168. Heerma van Voss, M.R., P.J. van Diest, and V. Raman, *Targeting RNA helicases in cancer: The translation trap.* Biochim Biophys Acta, 2017. **1868**(2): p. 510-520.
169. Rubio, C.A., et al., *Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation.* Genome Biol, 2014. **15**(10): p. 476.
170. Cencic, R., et al., *Antitumor activity and mechanism of action of the cyclopenta[b]benzofuran, silvestrol.* PLoS One, 2009. **4**(4): p. e5223.
171. Wolfe, A.L., et al., *RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer.* Nature, 2014. **513**(7516): p. 65-70.
172. Yang, L., et al., *A double tyrosine phosphorylation of P68 RNA helicase confers resistance to TRAIL-induced apoptosis.* Oncogene, 2007. **26**(41): p. 6082-92.

173. Yang, L., et al., *Phosphorylation of p68 RNA helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin D1 and c-Myc expression.* J Biol Chem, 2007. **282**(23): p. 16811-9.
174. Carter, C.L., et al., *Phosphorylated p68 RNA helicase activates Snail1 transcription by promoting HDAC1 dissociation from the Snail1 promoter.* Oncogene, 2010. **29**(39): p. 5427-36.
175. Yang, L., C. Lin, and Z.R. Liu, *Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation.* Mol Cancer Res, 2005. **3**(6): p. 355-63.
176. Kost, G.C., et al., *A Novel Anti-Cancer Agent, 1-(3,5-Dimethoxyphenyl)-4-[(6-Fluoro-2-Methoxyquinoxalin-3-yl)Aminocarbonyl] Piperazine (RX-5902), Interferes With beta-Catenin Function Through Y593 Phospho-p68 RNA Helicase.* J Cell Biochem, 2015. **116**(8): p. 1595-601.
177. Lee, Y., R. Mazhari, and D.J. Kim, *The anticancer effects of supinoxin (RX-5902) in renal cell cancer.* Journal of Clinical Oncology, 2016. **34**(2).
178. Eckhardt, S.G., et al., *A phase 1 study of RX-5902, an oral agent targeting phosphorylated p68, to treat subjects with advanced solid tumors.* Journal of Clinical Oncology, 2015. **33**(15).
179. Lee, Y., R. Mazhari, and D.J. Kim, *The anticancer effects of supinoxin (RX-5902) in pancreatic carcinoma.* Journal of Clinical Oncology, 2016. **34**(4).
180. Rajendran, R.R., et al., *Regulation of nuclear receptor transcriptional activity by a novel DEAD box RNA helicase (DP97).* J Biol Chem, 2003. **278**(7): p. 4628-38.
181. Kanno, Y., et al., *DP97, a DEAD box DNA/RNA helicase, is a target gene-selective co-regulator of the constitutive androstane receptor.* Biochem Biophys Res Commun, 2012. **426**(1): p. 38-42.
182. Tafforeau, L., et al., *The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors.* Mol Cell, 2013. **51**(4): p. 539-51.
183. Milek, M., et al., *DDX54 regulates transcriptome dynamics during DNA damage response.* Genome Res, 2017. **27**(8): p. 1344-1359.
184. Tarn, W.Y. and T.H. Chang, *The current understanding of Ded1p/DDX3 homologs from yeast to human.* RNA Biol, 2009. **6**(1): p. 17-20.
185. Chang, T.C. and W.S. Liu, *The molecular evolution of PL10 homologs.* BMC Evol Biol, 2010. **10**: p. 127.
186. Sharma, D. and E. Jankowsky, *The Ded1/DDX3 subfamily of DEAD-box RNA helicases.* Crit Rev Biochem Mol Biol, 2014. **49**(4): p. 343-60.
187. Kim, Y.S., et al., *Gene structure of the human DDX3 and chromosome mapping of its related sequences.* Mol Cells, 2001. **12**(2): p. 209-14.
188. Ditton, H.J., et al., *The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control.* Hum Mol Genet, 2004. **13**(19): p. 2333-41.
189. Kotov, A.A., et al., *Progress in understanding the molecular functions of DDX3Y (DBY) in male germ cell development and maintenance.* Biosci Trends, 2017. **11**(1): p. 46-53.

190. Hogbom, M., et al., *Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP.* J Mol Biol, 2007. **372**(1): p. 150-9.
191. Floor, S.N., et al., *Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3.* J Biol Chem, 2016. **291**(5): p. 2412-21.
192. Putnam, A.A., et al., *Division of Labor in an Oligomer of the DEAD-Box RNA Helicase Ded1p.* Mol Cell, 2015. **59**(4): p. 541-52.
193. Yedavalli, V.S., et al., *Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function.* Cell, 2004. **119**(3): p. 381-92.
194. Hilliker, A., et al., *The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex.* Mol Cell, 2011. **43**(6): p. 962-72.
195. Zhou, Z., et al., *Comprehensive proteomic analysis of the human spliceosome.* Nature, 2002. **419**(6903): p. 182-5.
196. Merz, C., et al., *Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment.* RNA, 2007. **13**(1): p. 116-28.
197. Lai, M.C., Y.H. Lee, and W.Y. Tarn, *The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control.* Mol Biol Cell, 2008. **19**(9): p. 3847-58.
198. Soto-Rifo, R., et al., *DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs.* EMBO J, 2012. **31**(18): p. 3745-56.
199. Lee, C.S., et al., *Human DDX3 functions in translation and interacts with the translation initiation factor eIF3.* Nucleic Acids Res, 2008. **36**(14): p. 4708-18.
200. Shih, J.W., et al., *Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response.* Biochem J, 2012. **441**(1): p. 119-29.
201. Shih, J.W., et al., *Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein.* Oncogene, 2008. **27**(5): p. 700-14.
202. Lai, M.C., et al., *DDX3 regulates cell growth through translational control of cyclin E1.* Mol Cell Biol, 2010. **30**(22): p. 5444-53.
203. Valentin-Vega, Y.A., et al., *Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation.* Sci Rep, 2016. **6**: p. 25996.
204. Ramathal, C., et al., *DDX3Y gene rescue of a Y chromosome AZFa deletion restores germ cell formation and transcriptional programs.* Sci Rep, 2015. **5**: p. 15041.
205. Gueler, B., et al., *AZFa protein DDX3Y is differentially expressed in human male germ cells during development and in testicular tumours: new evidence for phenotypic plasticity of germ cells.* Hum Reprod, 2012. **27**(6): p. 1547-55.
206. Zhao, L., et al., *Multifunctional DDX3: dual roles in various cancer development and its related signaling pathways.* Am J Cancer Res, 2016. **6**(2): p. 387-402.
207. Bol, G.M., M. Xie, and V. Raman, *DDX3, a potential target for cancer treatment.* Mol Cancer, 2015. **14**: p. 188.

208. Stransky, N., et al., *The mutational landscape of head and neck squamous cell carcinoma.* Science, 2011. **333**(6046): p. 1157-60.
209. Wu, D.W., et al., *Reduced p21(WAF1/CIP1) via alteration of p53-DDX3 pathway is associated with poor relapse-free survival in early-stage human papillomavirus-associated lung cancer.* Clin Cancer Res, 2011. **17**(7): p. 1895-905.
210. Su, C.Y., et al., *DDX3 as a strongest prognosis marker and its downregulation promotes metastasis in colorectal cancer.* Oncotarget, 2015. **6**(21): p. 18602-12.
211. Chang, P.C., et al., *DDX3, a DEAD box RNA helicase, is deregulated in hepatitis virus-associated hepatocellular carcinoma and is involved in cell growth control.* Oncogene, 2006. **25**(14): p. 1991-2003.
212. Chao, C.H., et al., *DDX3, a DEAD box RNA helicase with tumor growth-suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor.* Cancer Res, 2006. **66**(13): p. 6579-88.
213. Northcott, P.A., et al., *Medulloblastomics: the end of the beginning.* Nat Rev Cancer, 2012. **12**(12): p. 818-34.
214. Kool, M., et al., *Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition.* Cancer Cell, 2014. **25**(3): p. 393-405.
215. Robinson, G., et al., *Novel mutations target distinct subgroups of medulloblastoma.* Nature, 2012. **488**(7409): p. 43-8.
216. Bol, G.M., et al., *Targeting DDX3 with a small molecule inhibitor for lung cancer therapy.* EMBO Mol Med, 2015. **7**(5): p. 648-69.
217. Heerma van Voss, M.R., et al., *Identification of the DEAD box RNA helicase DDX3 as a therapeutic target in colorectal cancer.* Oncotarget, 2015. **6**(29): p. 28312-26.
218. Wilky, B.A., et al., *RNA helicase DDX3: a novel therapeutic target in Ewing sarcoma.* Oncogene, 2016. **35**(20): p. 2574-83.
219. Dai, X., et al., *Breast cancer intrinsic subtype classification, clinical use and future trends.* Am J Cancer Res, 2015. **5**(10): p. 2929-43.
220. Botlagunta, M., et al., *Oncogenic role of DDX3 in breast cancer biogenesis.* Oncogene, 2008. **27**(28): p. 3912-22.
221. Xie, M., et al., *NZ51, a ring-expanded nucleoside analog, inhibits motility and viability of breast cancer cells by targeting the RNA helicase DDX3.* Oncotarget, 2015. **6**(30): p. 29901-13.
222. Botlagunta, M., et al., *Expression of DDX3 is directly modulated by hypoxia inducible factor-1 alpha in breast epithelial cells.* PLoS One, 2011. **6**(3): p. e17563.
223. Bol, G.M., et al., *Expression of the RNA helicase DDX3 and the hypoxia response in breast cancer.* PLoS One, 2013. **8**(5): p. e63548.
224. Sun, M., et al., *DDX3 regulates DNA damage-induced apoptosis and p53 stabilization.* Biochim Biophys Acta, 2013. **1833**(6): p. 1489-97.
225. Heerma van Voss, M.R., et al., *The prognostic effect of DDX3 upregulation in distant breast cancer metastases.* Clin Exp Metastasis, 2017. **34**(1): p. 85-92.

226. Heerma van Voss, M.R., et al., *Nuclear DDX3 expression predicts poor outcome in colorectal and breast cancer.* Onco Targets Ther, 2017. **10**: p. 3501-3513.
227. Pui, C.H., L.L. Robison, and A.T. Look, *Acute lymphoblastic leukaemia.* Lancet, 2008. **371**(9617): p. 1030-43.
228. Estey, E. and H. Dohner, *Acute myeloid leukaemia.* Lancet, 2006. **368**(9550): p. 1894-907.
229. Chiorazzi, N., K.R. Rai, and M. Ferrarini, *Chronic lymphocytic leukemia.* N Engl J Med, 2005. **352**(8): p. 804-15.
230. Jabbour, E. and H. Kantarjian, *Chronic myeloid leukemia: 2014 update on diagnosis, monitoring, and management.* Am J Hematol, 2014. **89**(5): p. 547-56.
231. Ojha, J., et al., *Identification of recurrent truncated DDX3X mutations in chronic lymphocytic leukaemia.* Br J Haematol, 2015. **169**(3): p. 445-8.
232. Jiang, L., et al., *Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma.* Nat Genet, 2015. **47**(9): p. 1061-6.
233. Brandimarte, L., et al., *DDX3X-MLLT10 fusion in adults with NOTCH1 positive T-cell acute lymphoblastic leukemia.* Haematologica, 2014. **99**(5): p. 64-6.
234. Brandimarte, L., et al., *New MLLT10 gene recombinations in pediatric T-acute lymphoblastic leukemia.* Blood, 2013. **121**(25): p. 5064-7.
235. Tzelepis, K., et al., *A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia.* Cell Rep, 2016. **17**(4): p. 1193-1205.
236. Suzuki, K., et al., *READ: A two minute cell fractionation method.* BMC Res Notes, 2010. **3**: p. 294.
237. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
238. Lawrence, M., et al., *Software for computing and annotating genomic ranges.* PLoS Comput Biol, 2013. **9**(8): p. e1003118.
239. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
240. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
241. Aken, B.L., et al., *Ensembl 2017.* Nucleic Acids Res, 2017. **45**(D1): p. D635-D642.
242. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
243. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
244. Huppertz, I., et al., *iCLIP: protein-RNA interactions at nucleotide resolution.* Methods, 2014. **65**(3): p. 274-87.
245. Konig, J., et al., *Protein-RNA interactions: new genomic technologies and perspectives.* Nat Rev Genet, 2012. **13**(2): p. 77-83.
246. Sasaki, Y.T.F., et al., *MEN epsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles.* Proceedings of the National

Academy of Sciences of the United States of America, 2009. **106**(8): p. 2525-2530.
247. Hartman, T.R., et al., *RNA helicase A is necessary for translation of selected messenger RNAs.* Nat Struct Mol Biol, 2006. **13**(6): p. 509-16.
248. Chen, X., et al., *Kruppel-like factor 4 (gut-enriched Kruppel-like factor) inhibits cell proliferation by blocking G1/S progression of the cell cycle.* J Biol Chem, 2001. **276**(32): p. 30423-8.
249. Shields, J.M., R.J. Christy, and V.W. Yang, *Identification and characterization of a gene encoding a gut-enriched Kruppel-like factor expressed during growth arrest.* J Biol Chem, 1996. **271**(33): p. 20009-17.
250. McConnell, B.B. and V.W. Yang, *Mammalian Kruppel-like factors in health and diseases.* Physiol Rev, 2010. **90**(4): p. 1337-81.
251. Tetreault, M.P., Y. Yang, and J.P. Katz, *Kruppel-like factors in cancer.* Nat Rev Cancer, 2013. **13**(10): p. 701-13.
252. Ghaleb, A.M. and V.W. Yang, *Kruppel-like factor 4 (KLF4): What we currently know.* Gene, 2017. **611**: p. 27-37.
253. Yam, C.H., T.K. Fung, and R.Y. Poon, *Cyclin A in cell cycle control and cancer.* Cell Mol Life Sci, 2002. **59**(8): p. 1317-26.
254. Blanchard, J.M., *Cyclin A2 transcriptional regulation: modulation of cell cycle control at the G1/S transition by peripheral cues.* Biochem Pharmacol, 2000. **60**(8): p. 1179-84.
255. Aleem, E., C. Berthet, and P. Kaldis, *Cdk2 as a master of S phase entry: fact or fake?* Cell Cycle, 2004. **3**(1): p. 35-7.
256. Otto, T. and P. Sicinski, *Cell cycle proteins as promising targets in cancer therapy.* Nat Rev Cancer, 2017. **17**(2): p. 93-115.
257. Abbas, T. and A. Dutta, *p21 in cancer: intricate networks and multiple activities.* Nat Rev Cancer, 2009. **9**(6): p. 400-14.
258. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
259. Zhou, Y., et al., *High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells.* Nature, 2014. **509**(7501): p. 487-91.
260. Kiessling, M.K., et al., *Identification of oncogenic driver mutations by genome-wide CRISPR-Cas9 dropout screening.* BMC Genomics, 2016. **17**(1): p. 723.
261. Shi, J., et al., *Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains.* Nat Biotechnol, 2015. **33**(6): p. 661-7.
262. Zuber, J., et al., *Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi.* Nat Biotechnol, 2011. **29**(1): p. 79-83.
263. Quentmeier, H., et al., *Cell line OCI/AML3 bears exon-12 NPM gene mutation-A and cytoplasmic expression of nucleophosmin.* Leukemia, 2005. **19**(10): p. 1760-7.
264. Box, J.K., et al., *Nucleophosmin: from structure and function to disease development.* BMC Mol Biol, 2016. **17**(1): p. 19.
265. Grisendi, S., et al., *Nucleophosmin and cancer.* Nat Rev Cancer, 2006. **6**(7): p. 493-505.
266. Heath, E.M., et al., *Biological and clinical consequences of NPM1 mutations in AML.* Leukemia, 2017. **31**(4): p. 798-807.

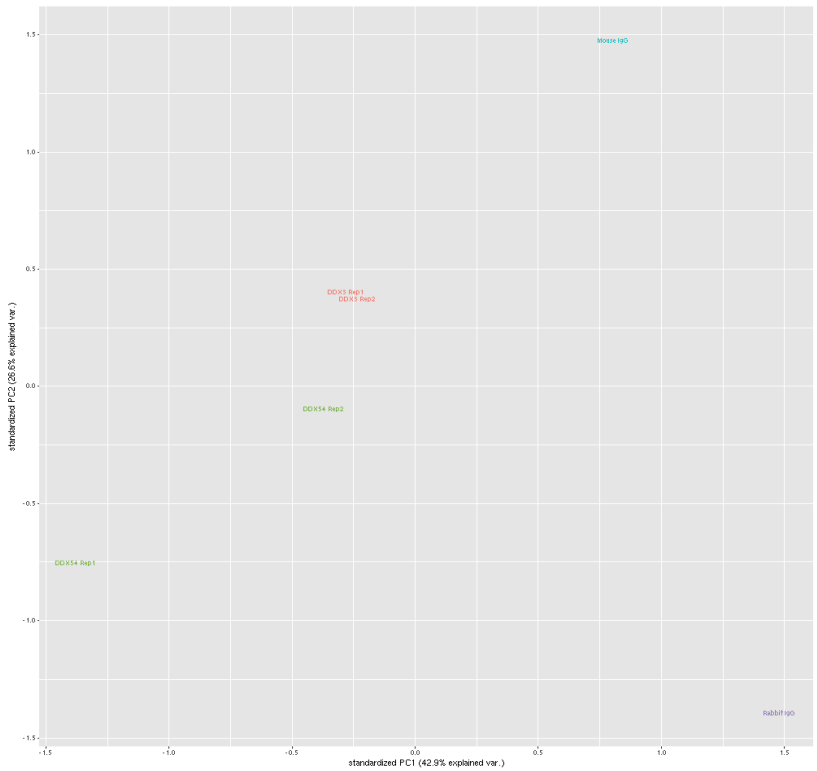
267. Cheng, K., et al., *The cytoplasmic NPM mutant induces myeloproliferation in a transgenic mouse model.* Blood, 2010. **115**(16): p. 3341-5.
268. Adriaens, C., et al., *p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity.* Nat Med, 2016. **22**(8): p. 861-8.
269. Dickey, J.S., et al., *Intercellular communication of cellular stress monitored by gamma-H2AX induction.* Carcinogenesis, 2009. **30**(10): p. 1686-95.
270. Seker, H., et al., *UV-C-induced DNA damage leads to p53-dependent nuclear trafficking of PML.* Oncogene, 2003. **22**(11): p. 1620-8.
271. Karin, M., Z. Liu, and E. Zandi, *AP-1 function and regulation.* Curr Opin Cell Biol, 1997. **9**(2): p. 240-6.
272. Millena, A.C., B.T. Vo, and S.A. Khan, *JunD Is Required for Proliferation of Prostate Cancer Cells and Plays a Role in Transforming Growth Factor-beta (TGF-beta)-induced Inhibition of Cell Proliferation.* J Biol Chem, 2016. **291**(34): p. 17964-76.
273. Hernandez, J.M., et al., *Multiple facets of junD gene expression are atypical among AP-1 family members.* Oncogene, 2008. **27**(35): p. 4757-67.
274. Kondaskar, A., et al., *Novel, Broad Spectrum Anti-Cancer Agents Containing the Tricyclic 5:7:5-Fused Diimidazodiazepine Ring System.* ACS Med Chem Lett, 2010. **2**(3): p. 252-256.
275. Xie, M., et al., *RK-33 Radiosensitizes Prostate Cancer Cells by Blocking the RNA Helicase DDX3.* Cancer Res, 2016. **76**(21): p. 6340-6350.
276. Yedavalli, V.S., et al., *Ring expanded nucleoside analogues inhibit RNA helicase and intracellular human immunodeficiency virus type 1 replication.* J Med Chem, 2008. **51**(16): p. 5043-51.
277. Maga, G., et al., *Pharmacophore modeling and molecular docking led to the discovery of inhibitors of human immunodeficiency virus-1 replication targeting the human cellular aspartic acid-glutamic acid-alanine-aspartic acid box polypeptide 3.* J Med Chem, 2008. **51**(21): p. 6635-8.
278. Brai, A., et al., *Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents.* Proc Natl Acad Sci U S A, 2016. **113**(19): p. 5388-93.
279. Celik, H., et al., *Ezrin Binds to DEAD-Box RNA Helicase DDX3 and Regulates Its Function and Protein Level.* Mol Cell Biol, 2015. **35**(18): p. 3145-62.
280. Samal, S.K., et al., *Ketorolac salt is a newly discovered DDX3 inhibitor to treat oral cancer.* Sci Rep, 2015. **5**: p. 9982.

Appendices

A

Treatment	Total	Trimmed	No rDNA	Unique	Multihits	Filtered
DDX3 Rep1	7,198,334	6,112,711	5,456,099	2,014,556	452,399	2,014,556
DDX3 Rep2	8,298,062	7,717,846	6,662,403	3,606,680	686,105	3,606,680
DDX54 Rep1	9,169,546	8,060,557	7,752,847	2,447,920	555,865	2,447,920
DDX54 Rep2	14,038,767	12,564,711	11,935,399	5,353,547	1,186,656	5,353,547
Mouse IgG	152,317	106,857	97,792	18,270	5,040	18,270
Rabbit IgG	290,203	111,714	108,714	11,173	2,501	11,173

B



Appendix Figure 1- DDX3X and DDX54 iCLIP in MCF7 cells :reads and quality control analysis.

(A)Number of reads and proportion of reads following mapping and filtering of all biological replicates. (B) Quality Control: Principal Component Analysis of iCLIP biological replicates. Principal Component Analysis was done with the R function prcomp (with scale parameter set to true) on the transposed and log-transformed

matrix of normalised counts. The PCA biplot was realised with the R package "ggbiplot".

iCLIP raw data are available at the following link
http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6230
Username:Reviewer_E-MTAB-6230

Password: edchpurB

Treatment	Total	Trimmed	No rDNA	Unique	Multihits	Filtered
DDX3 Rep1	21,284,553	21,012,195		17,402,260	2,362,634	3,845,557
DDX3 Rep2	23,546,156	23,543,024		19,165,881	2,629,227	6,667,672
DDX3 Rep3	19,928,048	19,926,180		16,020,171	2,226,968	1,856,438
DDX54 Rep1	30,803,759	30,737,062		25,509,516	3,492,738	6,277,032
DDX54 Rep2	22,689,466	22,426,241		18,713,684	2,423,272	6,807,400
DDX54 Rep3	24,122,740	24,113,225		19,948,916	2,831,500	10,647,970
ER Rep1	15,468,087	15,466,999		12,688,298	1,971,911	4,329,449
ER Rep2	27,157,835	27,093,600		22,034,557	3,567,278	14,888,209
ER Rep3	29,173,483	29,089,524		23,683,625	3,803,242	16,859,047
IgG Rep1	13,056,341	12,646,157		10,258,124	1,393,843	6,653,846
IgG Rep2	2,632,617	2,630,624		2,167,319	293,560	389,110
IgG Rep3	8,221,927	8,027,596		6,588,545	921,418	478,242
Input Rep1	29,476,435	29,376,125		25,048,024	2,768,650	7,589,118
Input Rep2	21,608,354	21,604,508		18,104,471	2,404,049	13,473,714
Input Rep3	30,253,873	30,203,900		24,705,830	3,965,329	21,007,279
H3K4me3 Rep1	18,289,254	18,276,611		15,426,818	1,805,447	10,658,837
H3K4me3 Rep2	17,755,110	17,727,699		13,651,368	2,923,217	11,565,206
H3K4me3 Rep3	19,744,094	19,742,913		15,493,966	3,010,066	10,033,978

Appendix Figure 2-ERα, DDX3X, DDX54 and H3K4me3 ChIP-seq in MCF7 cells: reads.

Number of reads and proportion of reads following mapping and filtering of all biological replicates.

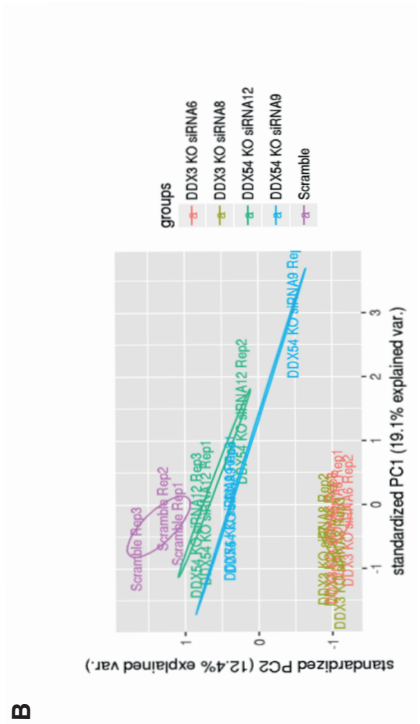
ChIP-seq raw data are available at the following link
http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6211
Username:Reviewer_E-MTAB-6211

Password: kk33morL

A

Treatment	Total	Trimmed	No rDNA	Unique	Multihits	Filtered
Scramble siRNA Rep1	19,854,506/19,854,506	19,851,231/19,851,231	14,640,478/14,640,478	9,335,156/9,335,156	4,189,494/4,189,494	9,335,156/9,335,156
#6 DDX3X siRNA Rep1	25,605,615/25,605,615	25,601,384/25,601,384	19,559,912/19,559,912	12,320,218/12,320,218	5,948,627/5,948,627	12,320,218/12,320,218
#8 DDX3X siRNA Rep1	15,870,518/15,870,518	15,868,902/15,868,902	12,718,353/12,718,353	8,000,652/8,000,652	3,837,130/3,837,130	8,000,652/8,000,652
#9 DDX54 siRNA Rep1	22,111,165/22,111,165	22,106,673/22,106,673	15,633,672/15,633,672	10,465,860/10,465,860	3,997,806/3,997,806	10,465,860/10,465,860
#12 DDX54 siRNA Rep1	21,921,324/21,921,324	21,919,562/21,919,562	15,248,574/15,248,574	9,674,345/9,674,345	4,410,435/4,410,435	9,674,345/9,674,345
Scramble siRNA Rep2	34,539,606/34,539,606	34,535,274/34,535,274	29,526,554/29,526,554	18,530,980/18,530,980	9,267,326/9,267,326	18,530,980/18,530,980
#6 DDX3X siRNA Rep2	17,431,543/17,431,543	17,429,511/17,429,511	12,863,185/12,863,185	8,295,750/8,295,750	3,517,562/3,517,562	8,295,750/8,295,750
#8 DDX3X siRNA Rep2	19,021,353/19,021,353	19,017,949/19,017,949	14,241,686/14,241,686	9,231,954/9,231,954	3,924,432/3,924,432	9,231,954/9,231,954
#9 DDX54 siRNA Rep2	25,375,359/25,375,359	25,372,201/25,372,201	15,136,739/15,136,739	10,823,233/10,823,233	3,404,785/3,404,785	10,823,233/10,823,233
#12 DDX54 siRNA Rep2	22,671,055/22,671,055	22,668,092/22,668,092	13,952,768/13,952,768	9,535,033/9,535,033	3,436,574/3,436,574	9,535,033/9,535,033
Scramble siRNA Rep3	33,968,223/33,968,223	33,965,488/33,965,488	28,916,829/28,916,829	20,477,164/20,477,164	6,413,023/6,413,023	20,477,164/20,477,164
#6 DDX3X siRNA Rep3	28,590,010/28,590,010	28,588,223/28,588,223	22,499,006/22,499,006	14,540,190/14,540,190	6,550,212/6,550,212	14,540,190/14,540,190
#8 DDX3X siRNA Rep3	25,522,132/25,522,132	25,518,239/25,518,239	21,257,746/21,257,746	13,511,624/13,511,624	6,304,249/6,304,249	13,511,624/13,511,624
#9 DDX54 siRNA Rep3	21,391,121/21,391,121	21,387,970/21,387,970	16,412,028/16,412,028	11,220,034/11,220,034	4,224,840/4,224,840	11,220,034/11,220,034
#12 DDX54 siRNA Rep3	18,125,699/18,125,699	18,122,311/18,122,311	13,898,939/13,898,939	9,251,533/9,251,533	3,787,740/3,787,740	9,251,533/9,251,533

B

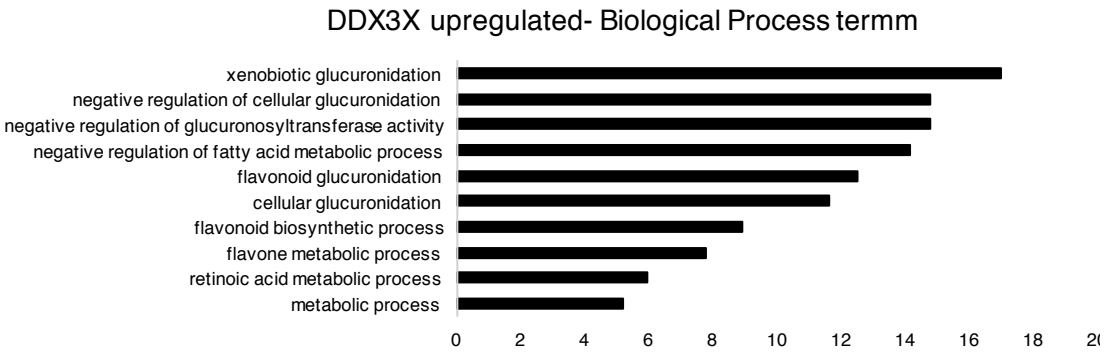


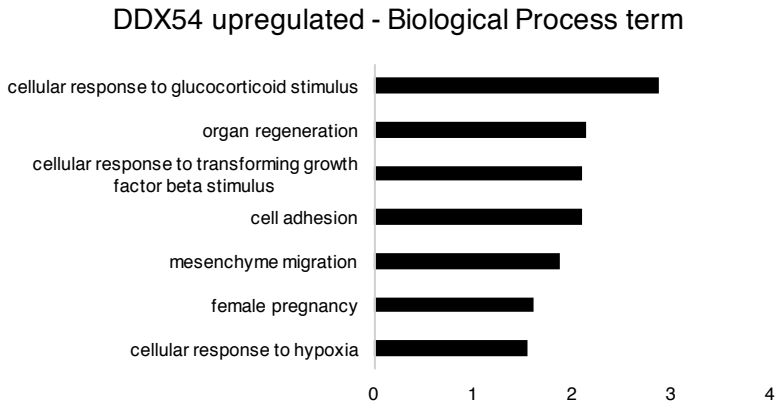
Appendix Figure 3- MCF7 RNA-seq upon DDX3X and DDX54 knock down: reads and quality control analysis.

(A)Number of reads and proportion of reads following mapping and filtering of all biological replicates. (B) Quality Control: Principal Component Analysis of RNA-seq biological replicates. Principal Component Analysis was done with the R function prcomp (with scale parameter set to true) on the transposed and log-transformed matrix of normalised counts. The PCA biplot was realised with the R package "ggbiplot".

RNA-seq raw data are available at the following link
https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6235
Username:Reviewer_E-MTAB-6235

Password: C73ooofp



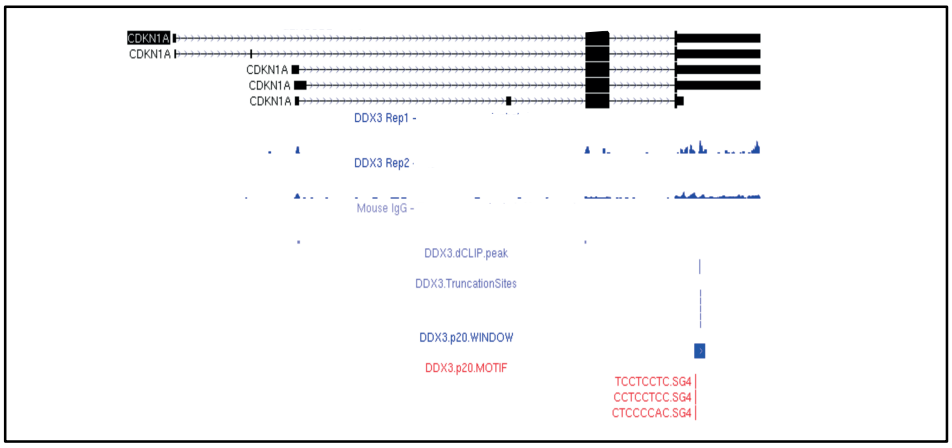


Appendix Figure 4 Up-regulated genes in MCF7 cells upon DDX3X and DDX54 knock down.

Upregulated genes DAVID GO terms in MCF7 cells knockdown for DDX3X and DDX54. The GO enrichment of differentially expressed genes was obtained using the DAVID functional annotation tool with default settings [PMID: 19131956]. The p-values were calculated using the "default" method of DAVID. X-axis shows the GO terms related to cellular components. Y-axis shows –log10 p-values of each GO terms.

Appendix Figure 5 iCLIP-RNA-seq intersection for DDX54.

Intersection of gene expression changes upon DDX54 knock down and DDX54 iCLIP targets in MCF7 cells.
Venn diagrams show the intersection between DDX54 iCLIP targets and genes upregulated (left) or downregulated (right) upon DDX54 knockdown detected in RNAseq.



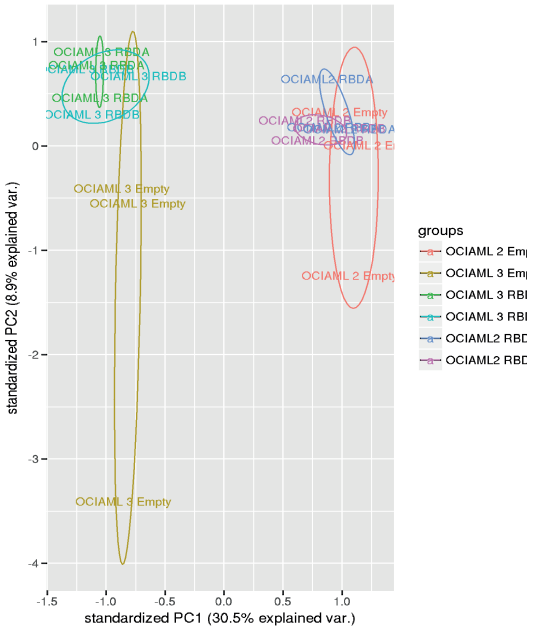
Appendix Figure 6- DDX3X binding motif on *CDKN1A* mRNA.

Visualization of the USCS custom tracks for DDX3X iCLIP truncation sites and motif localization on *CDKN1A (p21)* sequence.

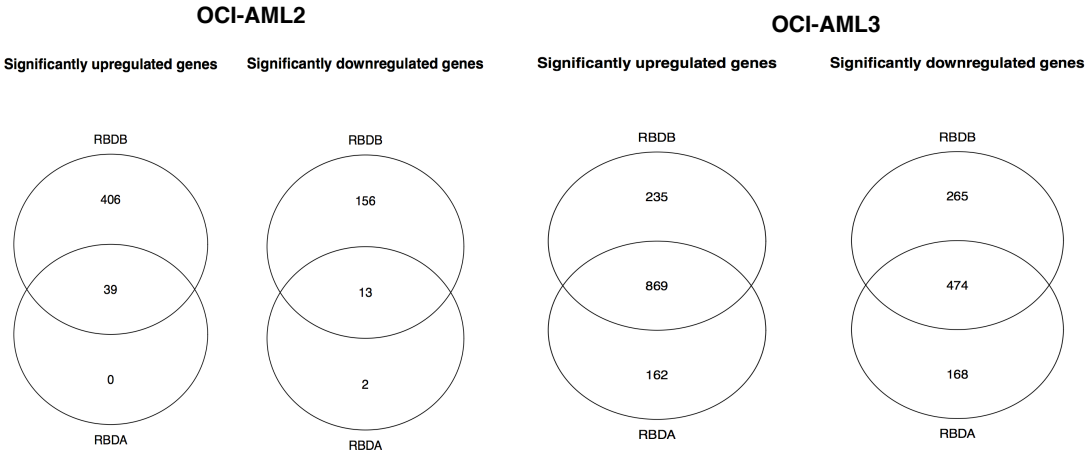
A

Sample	Reads
OCIAML 2 Empty REP 2	6336375
OCIAML 2 Empty REP1	19837407
OCIAML2 RBDA REP2	14964387
OCIAML2 RBDA REP1	15811992
OCIAML2 RBDB REP1	20268753
OCIAML 3 Empty REP1	26813844
OCIAML 3 RBDA REP1	15806384
OCIAML2 RBDB REP2	17333318
OCIAML 3 Empty REP2	4439568
OCIAML 3 RBDA REP2	19871471
OCIAML 3 RBDB REP2	15504811
OCIAML 3 RBDB REP1	11671363
OCIAML 2 Empty REP3	15700809
OCIAML2 RBDA REP3	33805347
OCIAML2 RBDB REP3	15857681
OCIAML 3 Empty REP3	16273186
OCIAML 3 RBDB REP3	10746980
OCIAML 3 RBDA REP3	19347511

B



C

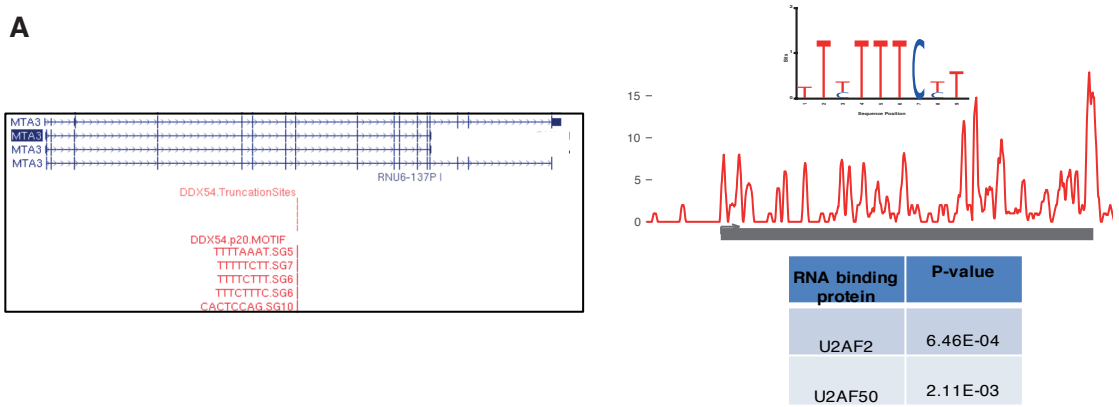


Appendix Figure 7- OCI-AML2/OCI-AML3 RNAseq RNA-seq upon DDX3X RNA binding domain (A and B) disruption: reads, quality control analysis and gene expression changes.

(A)Number of reads and proportion of reads following mapping and filtering of all biological replicates. (B) Quality Control: Principal Component Analysis. Principal Component Analysis was done with the R function prcomp (with scale parameter set to true) on the transposed and log-transformed matrix of normalised counts. The PCA biplot was realised with the R package "ggbiplot". (C) Venn diagrams sowing up- and down- regulated genes upon CRISPR/Cas9-mediated targeting of RNA binding domain A (RBDA) and B (RBDB) of DDX3X in either OCI-AML2 or OCI-AML3 cells.

RNA-seq raw data are available at the following link
http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6241
Username:Reviewer_E-MTAB-6241

Password: qqsrrhic



RNA binding protein	P-value
U2AF2	6.46E-04
U2AF50	2.11E-03

Appendix Figure 8-DDX54 binding motifs on *MTA3* mRNA and *NEAT1* lncRNA.

A.Visualization of the USCS custom tracks for DDX54 iCLIP truncation sites and motif localization on *MTA3 (p21)* sequence. Density plot of the motifs localisation within a scaled gene body. The scaled gene body represents three indicative regions: 5'-end, middle, and 3'-end. Y axis indicates the frequency (actual count) of the motif; X axis = scaled gene body localization. Consensus motifs for the clustered 8-mer(s) in each phylogenetic group were found with MATLAB function *seqlogo* (Section 2.13.5). The sequence logo of the phylogenetic group consensus motif is shown (Y axis = Bits; X axis =sequence position). The motif is matched

to known RNA binding motif. Table shows the known RNA binding proteins that recognize a similar sequence motif and relative p-values. (B). Visualization of the USCS custom tracks for DDX54 iCLIP truncation sites and motif localization on *NEAT1 lncRNA* sequence.